

**Deciding the Fast & Frugal Way on the Application of
Pharmacodiagnostic Tests in Cancer Care?**

**A Comparative Study of Oncologists', Pathologists', and Cancer
Patients' Decision Making in Germany and the USA**

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. rer. nat.)

im Fach Psychologie

Eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät II

Humboldt-Universität zu Berlin

Von:

Dipl.-Psych. Odette Wegwarth

11.06.1973, Salzwedel

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Christoph Marksches

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II

Prof. Dr. Wolfgang Coy

Gutachter:

1. Prof. Gerd Gigerenzer
2. Prof. Wolfgang Scholl
3. Prof. Ulrich Hoffrage

Tag der Verteidigung: 20.04.2007

Acknowledgement

For the outstanding scientific contributions, excellent discussions, exceptional advising, great inspirations, substantial work and generous support, heavy programming, nice designing, helpful comments on earlier drafts of this work, important proofreading, patience, friendship, trust, and love I would like to thank:

Gerd Gigerenzer – Steffi Kurzenhäuser – Wolfgang Scholl – Ulrich Hoffrage – Robert W.
Day – Gregor Lämmel – Juliet Conlin – Henrik Olsson – Jörg Rieskamp – Wolfgang
Gaissmaier – Ulrich Klocke – Anita Todd – Christel Fraser – The ABC Group – The HU
Colloquium



Epigenomics AG – Torsten Horns – Patricia Jaecklin – Fabian Model – Dennis Bieling –
Robert Sander – Carolina Haefliger – Sabine Maier – Achim Plum



699 mainly unknown participants of the studies



Prof. K. Possinger – Dr. Th. Decker – Dr. R. Pauli – Dr. Sabine von Österreich – Dr. F.
Strohbach – Prof. A. Tannapfel

Deutsche Krebsgesellschaft eV – Gilda's Club Worldwide – Komen Breast Cancer
Foundation & C.J. – Mamazone – Pink – Inka

Städtische Klinikum Magdeburg – Online Zeitung „Krebs & Mensch“ – Krebs-Kompass –
Deutsche Gesellschaft für Pathologie – Berufsverband Deutscher Pathologen
Arbeitsgemeinschaft Deutscher Tumorzentren – Nord-Ostdeutsche Gesellschaft für
Gynäkologische Onkologie – Deutsches Netzwerk Evidenzbasierter Medizin



Anne – Claudia – Hanene – Tine – Anja – Arne



Daniel and my parents

This thesis could not have been accomplished without you.

1	<i>An Introduction</i>	15
1.1	Cancer	15
1.2	Cancer Treatments	16
1.3	Pharmacodiagnostic Testing	18
1.4	The Goal of the Thesis	21
2	<i>When Physicians Have to Decide: A Case of Complexity?</i>	23
2.1	Some Words About Certain Uncertainty in Medicine	24
2.2	Condition of Medical Decision-Making Tasks	25
2.2.1	Decisions on Tests: A case of Task Complexity or the Bite of Bayes	25
2.2.2	Do Pharmacodiagnostic Tests Have the Potential to Smart Task Complexity and Uncertainty?	29
2.2.3	Time Pressure	30
2.2.4	Quality of Information	30
2.2.5	Outcome Feedback	31
2.2.6	Cost of Health Care	31
2.2.7	Medical Guidelines	32
2.2.8	Last But Not Least: Patients' Meddling In	33
2.2.9	Unbounded Bounded?	34
2.2.10	Clearing Up the Sky by Offering Bounds	36
2.2.11	Heuristic = Bias?	36
2.2.12	Do It Fast, Do It Frugal!	38
3	<i>Study 1: Discovering Oncologists' Test Decision Making</i>	41
3.1	Pilot Study	41
3.1.1	Design	41
3.1.2	Procedure	41
3.2	Sample characteristics	43
3.2.1	Participants	43

3.2.2	Analytical Procedure	44
3.2.3	Results	44
3.3	Implications for the Main Study: Do Oncologists Have a Utility Function in Their Minds?	50
3.4	Dive Into the Bliss of Researching Human Judgment and Decision Making	52
3.5	Main Study	54
3.5.1	Introduction	54
3.5.2	Design	55
3.5.3	Procedure	59
3.5.4	Participants	61
3.5.5	Analysis Considerations	63
3.5.6	Results	70
3.6	Summary & Discussion	88
4	<i>Study 2: What Role Do The Pathologists Play?</i>	95
4.1	Pilot Study	95
4.1.1	Design	95
4.1.2	Procedure	95
4.1.3	Participants	97
4.1.4	Analytical Procedure	97
4.2	Sample characteristic	98
4.2.1	Results	98
4.3	Implications for the Main Study – Do Pathologists Juggle With Cues Weights They Only Feel?	102
4.4	Main study	104
4.4.1	Design	105
4.4.2	Procedure	108
4.4.3	Participants	110
4.4.4	Analysis Considerations	112
4.4.5	Results	116

4.5	Summary & Discussion	134
5	<i>The Idea of Shared Decision Making—Or: Does It Always Need Two to Tango?</i>	140
5.1	Is to Tango Desired in the Field of Cancer?	142
5.2	Pilot Study	145
5.2.1	Design	145
5.2.2	Procedure	147
5.2.3	Participants	147
5.2.4	Analytical Procedure	149
5.2.5	Results	149
5.3	Implication for the Main Study: Is There Anything to Tango Regarding Tests?	153
5.4	Main Study	155
5.4.1	Introduction	155
5.4.2	Design	156
5.4.3	Procedure	160
5.4.4	Participants	162
5.4.5	Analysis Considerations	166
5.4.6	Results	168
5.5	Summary & Discussion	185
6	<i>What Have We Learned (So Far)?: Final Discussion and Conclusion Remarks</i>	190
6.1	Model Concerns or the Data's New Clothes	190
6.2	Implications for Research in Medical Decision Making	192
6.3	The Issue With Hypothetical "Paper" Cases and Some Other Methodological Concerns	195
6.3.1	Some Words on the Matching Heuristic	197
6.4	Implications for the Health System	198
	<i>References</i>	200

List of Figures

<i>Figure 3.1: Example of a flowchart of Matching Heuristic searching through a maximum of two cues ($K = 2$ model).</i>	68
<i>Figure 3.2: Proportion of yes-choices per country for all oncologists (All_oncologists) as well as separated between oncologists coming from university clinics (Uni_oncologists), hospitals (Hospital_oncologists), and private practices (Practice_oncologists).</i>	71
<i>Figure. 3.3: Disagreement over all 18 cases (9 per version) of the German oncologist sample (mean: 95% CI) for all oncologists (All_oncologists) as well as separated between oncologists coming from university clinics (Uni_oncologists), hospitals (Hospital_oncologists), and private practices (Practice_oncologists)</i>	72
<i>Figure 3.4: Disagreement over all 18 cases (9 per version) of the US oncologist sample (mean: 95% CI) for all oncologists (All_oncologists) as well as separated between oncologists coming from university clinics (Uni_oncologists), hospitals (Hospital_oncologists), and private practices (Practice_oncologists).</i>	73
<i>Figure 3.5: Proportion of the Best Fitting Model for a Participant's overall Choice Behavior provided by either Franklin's rule, Dawes' rule, Matching Heuristic, or by two or all Models for the German and the US oncologists.</i>	74
<i>Figure 3.6: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German oncologists ($N = 111$).</i>	76
<i>Figure 3.7: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US oncologists ($N = 109$).</i>	76
<i>Figure 3.8: Proportion of the best generalizing model for a participant's overall choice behavior provided by either Franklin's rule, Dawes' rule, Matching Heuristic, or by two or all models for the German and the US oncologists.</i>	78
<i>Figure 3.9: Results of average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German oncologists ($N = 111$).</i>	79
<i>Figure 3.10: Results of average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US oncologists ($N = 109$).</i>	79

<i>Figure 3.11: Example for a German $K=1$ model (70.30% of all German choices) displaying the most frequently used cue.</i>	80
<i>Figure 3.12: Cue importance of Franklin's rule for the German as well as the US oncologist sample in the generalization setting.</i>	81
<i>Figure 3.13: Proportion of yes-choices (case wise) for recommended and nonrecommended cases per country (H1a).</i>	84
<i>Figure 3.14: Proportion of yes-choices (case wise) per country for nonrecommended cases either having the less or the high side effect cue value (H1b).</i>	85
<i>Figure 3.15: Proportion of the yes-choices (mean: 95% CI) per country of research-involved oncologists (Research_oncologists) vs. nonresearch involved oncologists (Non_rese_oncologists) for nonrecommended cases (three cases per questionnaire version).</i>	86
<i>Figure 3.16: Natural frequency tree for a hypothetical tropical disease having a prevalence of 30 percent and a hypothetical test with a sensitivity and specificity of 80 percent</i>	92
<i>Figure 4.1: Proportion of yes-choices per test alternative shown for all pathologists (overall) as well as separated for each country.</i>	118
<i>Figure 4.2: Disagreement (mean: 95% CI) over the 16 cases (8 per version) of the German pathologist sample for all pathologists (All_pathologists) as well as separated for pathologists coming from university clinics (Uni_pathologists), hospitals (Hospita_patyhologists), and private practices (Practic_pathologists).</i>	119
<i>Figure 4.3: Disagreement (mean: 95% CI) over the 16 cases (8 per version) of the US pathologist sample for all pathologists (All_pathologists) as well as separated for pathologists coming from university clinics (Uni_pathologists), hospitals (Hospita_pathologists), and private practices (Practic_pathologists).</i>	120
<i>Figure 4.4: Proportion of the best fitting model for a participant's overall choice behavior provided by either Franklin's rule, Dawes' rule, Take The Best, or by two or all models for the German and the US pathologists.</i>	121
<i>Figure 4.5: Results of the average fit (mean: 95% CI) for Franklin's rule, Dawes' rule, and Take The Best for the German pathologists.</i>	122

<i>Figure 4.6: Results of the average fit (mean: 95% CI) for Franklin's rule, Dawes' rule, and Take The Best for the US pathologists.</i>	123
<i>Figure 4.7: Proportion of the best generalizing model for a participant's overall choice behavior provided by either Franklin's rule, Dawes' rule, Take The Best, or by two or all models for the German and the US pathologists.</i>	124
<i>Figure 4.8: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Take The Best for the German pathologists (N = 93).</i>	125
<i>Figure 4.9: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Take The Best for the US pathologists (N= 108).</i>	126
<i>Figure 4.10: Cue importance of Franklin's rule for the generalization setting per country.</i>	127
<i>Figure 4.11: Range and mean for the proportion in opting for test A, given a higher positive value and an equal/lower positive value, respectively, than test B, per country.</i>	131
<i>Figure 4.12: Proportion in opting for a nonmicroscopic test per country depending on the location of work of the pathologists (H2a).</i>	132
<i>Figure 5.1: Proportion of yes-choices shown for all patients as well as separated for each country.</i>	169
<i>Figure 5.2: Disagreement (mean: 95% CI) over all 18 cases (9 per version) of the German patient sample as well as the US patient sample.</i>	169
<i>Figure 5.3: Proportion of the best fitting model for a participant's overall choice behavior—provided by either of Franklin's rule, Dawes' rule, Matching Heuristic or by two or all models for the German and the US patients.</i>	171
<i>Figure 5.4: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German patients (N = 116).</i>	172
<i>Figure 5.5: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US patients (N = 111).</i>	172
<i>Figure 5.6: Proportion of the best generalizing model for a participant's overall choice behavior—provided by either Franklin's rule, Dawes' rule, Matching Heuristic or by two or all models for the German and the US patients.</i>	175

<i>Figure 5.7: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German patients (N=116).</i>	175
<i>Figure 5.8: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US patients (N = 111).</i>	176
<i>Figure 5.9: Example of a K=1 model for a US patient displaying the most frequently used cue within this sample.</i>	178
<i>Figure 5.10: Proportion of yes-choices for recommended and nonrecommended cases per country (H1a).</i>	180
<i>Figure 5.11: Number of yes-choices per country for recommended cases (three per questionnaire version) and nonrecommended cases (three per questionnaire version) depending on the patient's reported relationship to the oncologist (H1b).</i>	181
<i>Figure 5.12: Proportion of yes-choices per country for nonrecommended tests depending on the patient's state of concern regarding the therapy's side effects.</i>	185

List of Tables

<i>Tab. 1: Different Test Models in the Field of Pharmacodiagnostic</i>	19
<i>Table 3.1: Catalogue of the Interview Questions for Oncologists</i>	42
<i>Table 3.2: Description of the Nineteen Participating Oncologists</i>	43
<i>Table 3.3: Cues Important for the Oncologists' Choice on Applying a Pharmacodiagnostic Test, Presented by Numbers of Mention</i>	49
<i>Table 3.4: Manipulated Cues, Their Levels and Distribution for Versions A and B of the Questionnaire</i>	57
<i>Table 3.5: Questions Concerning the Quality of the Questionnaire</i>	58
<i>Table 3.6: Questions and Their Values for Capturing Intragroup Differences and Demographics</i>	59
<i>Table 3.7: Demography and Intragroup-Related Values of the Main Study Oncologists</i>	62
<i>Table 3.8: Values of Reported Representativeness of the Questionnaire</i>	82
<i>Table 4.1: Catalogue of Interview Questions for Pathologists</i>	96
<i>Table 4.2: Description of the Fifteen Participating Pathologists</i>	98
<i>Table 4.3: Cues Important for the Pathologists' Choice on Applying a Pharmacodiagnostic Test, Presented by Numbers of Mention</i>	101
<i>Table 4.4: Cues, their Levels, and Distributions for Versions A and B of the Questionnaire for the Manipulated Test Alternative</i>	106
<i>Table 4.5: Cues and Their Levels for the Constant Test Alternative</i>	107
<i>Table 4.6: Questions and Their Possible Responses for Capturing Intragroup Differences and Demographics</i>	107
<i>Table 4.7: Demographics and Intragroup Related Values of Main Study Pathologists</i>	110
<i>Table 4.8: Dichotomized Cue Values of Former Polytomous Cues of the Pathologists' Main Study</i>	112

<i>Table 4.9: Values of Reported Representativeness of the Questionnaire</i>	129
<i>Table 4.10: Mean and Standard Deviation for Each Model Exhibit for German Pathologists Coming From University Clinics (1) As Well As for Pathologists Coming From Hospitals and Private Practices (2)</i>	133
<i>Table 4.11: Mean and Standard Deviation for Each Model Exhibit for US Pathologists Coming From University Clinics (1) As Well As for Pathologists Coming From Hospitals and Private Practices (2)</i>	133
<i>Table 5.1: Catalogue of the Interview Questions for Patients</i>	146
<i>Table 5.2: Description of the Seventeen Participating Pilot Study Patients</i>	148
<i>Table 5.3: Cues Important for Patients' Choice on Applying a Pharmacodiagnostic Test Presented by Number of Mentioning</i>	151
<i>Table 5.4: Manipulated Cues, Their Levels and Distribution for Version A and B of the Questionnaire</i>	158
<i>Table 5.5: Questions and Their Values for Capturing Intragroup Differences and Demographics</i>	158
<i>Table 5.6: Demography and Intragroup-Related Values of the Patients' Main Study</i>	164
<i>Note: Numbers do not always add up to the overall sample size, since not all questions had to be answered depending on the respective previous answer as well as due missing data</i>	166
<i>Table 5.7: Dichotomized Cue Values of Patients' Main Study</i>	167
<i>Table 5.8: Values of Reported Representativeness of the Patient Questionnaire</i>	179
<i>Table 5.9: Range of Proportion, Mean, and Standard Deviation of Yes-Choices Exhibited for Each Cost Values and Country</i>	183

1 An Introduction

1.1 Cancer

Being diagnosed with cancer commonly comes as a shock to a patient, and often leads to the patient entering a stage of crisis. For the majority of people, the diagnosis of cancer is synonymous with a death sentence. Looking at mortality figures might confirm these beliefs. Worldwide approximately 56 million people are diagnosed with cancer out of which every year 7 million people die, that is 12.5% of deaths worldwide. More than 11 million people are diagnosed with cancer every year. It is estimated that there will be 16 million new cases every year by 2020, and more than one in three of us will get some form of cancer at some point in our lives. Beyond what this disease means for an individual, the financial costs of cancer, alone for the USA in 2004, were approximately \$209.9 billion in total¹ (American Cancer Society, 2006).

In more medical terms, cancer is the uncontrolled growth and spread of cells that may affect almost any tissue of the body. In general, this broad disease group can be separated into solid tumors and hematological ones. The latter group refers to all kinds of cancer that are blood related, while the first group refers to all the cancer types that effect organs. Although there is already certain knowledge about promoters, for example, smoking and other lifestyle-related issues, explaining at least one third of all cancers existent which provides a solid base for recommendations for prevention, the main part of the variance related to the genesis of the cancer syndrome remains either unexplained or noninfluential by human beings' behavior.

This fact, combined with rather poor survival rates, is what makes the diagnosis of cancer so dreadful. The poor survival rates of most of the cancer syndromes are largely explained by less efficient, but often quite toxic, treatment options. Even if some of the most frequent cancer types may be curable by surgery, chemotherapy, or radiotherapy, to a reasonable extent when detected early, the overall chance of becoming cured by current available treatments is moderate at best. There are, of course, some advanced forms of treatment available that may produce a 5-year survival rate of 75% or more for certain types of

¹ The total cost figure includes a total of all health expenditures of \$74.0 billion, indirect morbidity costs (cost of lost productivity due to illness) of \$17.5 billion, as well as indirect mortality costs (cost of lost productivity due to premature death) of \$118.4 billion (American Cancer Society, 2006).

cancer, for example, cancer of the uterine corpus, breast, testis, and melanoma, if diagnosed in the adjuvant setting². By contrast, though, survival rates of cancer of the pancreas, liver, stomach, and lung are generally less than 15%, while it should be noted that the better as well as the worse survival rates are usually only achieved by administering more than one single treatment.

1.2 Cancer Treatments

Cancer treatments, such as chemotherapy, commonly entail moderate to even life-threatening bodily harm on patients, as they do not only affect mutated tissue and organs but also healthy ones. Generally, anticancer drugs affect cells that divide rapidly. In addition to cancer cells, these include blood cells, which fight infection, help the blood to clot, and carry oxygen to all parts of the body. When blood cells are affected, patients are more likely to contract infections, may bruise or bleed easily, and may feel unusually weak and very tired. Rapidly dividing cells in hair roots and cells that line the digestive tract may also be affected. As a result, chemotherapy side effects may include loss of hair, poor appetite, nausea and vomiting, diarrhea, or mouth and lip sores.

Certainly, chemotherapy side effects depend on many factors, including the size and location of the tumor, the treatment dose, and the part of body that is treated, as well as the patient's general health status. Comparable to any kind of treatments, side effects vary from person to person. Chemotherapy is usually given in so-called "cycles," which are a period of treatment followed by a period without treatment. Most side effects occur only gradually during the treatment phase, but vanish during the recovery periods between treatments, while some anticancer drugs can cause long-term side effects, such as the loss of fertility (the ability to produce children), or irreversible cardio-vascular damage. But adverse drug reactions are also an economical issue. A study carried out on 18,820 noncancer patients in England projected the annual cost of hospital admissions, due to adverse drug reaction, to the National Health System as £466 million (Pirmohamed et al., 2004). Lazarou, Pomeranz and Corey (1998) suggested in their study that adverse drug reactions are between the fourth and the sixth largest cause of death in US hospitals.

² Adjuvant setting (in medical literature referred to as Stages I to III) describes cancer syndromes, which have not developed metastases yet, and, therefore, are considered to be curable by treatment.

In one of the most promising turns in cancer treatment in years, a new generation of drugs is being approved that appear to prolong survival and are far less toxic than standard chemotherapy. While this has the potential to vastly improve the quality of life for patients, many of these treatments are extraordinarily expensive – predominately ones that target particularly cancerous cells but do not destroy healthy ones as current chemotherapies do. Many of these newer and so-called targeted therapies do not cure cancer, but rather keep it in check. The upshot is, however, that the cost of drugs is now becoming a critical element. While, for instance, until recently for colorectal cancer the standard treatment was Fluorouracil, also known as 5-FU, together with a vitamin called Leucovorin, that costs currently approximately \$500 for an average-sized individual, the five new targeted therapies available for advanced colorectal cancer, including Eloxatin, Erbitux, and Avastin, are closer to \$250,000 (Dockser Marcus, 2004). Each of these targeted therapies was launched with a so-called target test. The reason why these therapies come with a test is simple: No health insurance would reimburse their high cost if they were administered to every patient by default. Patient's eligibility to the therapy has to be proven by the test first. Therefore, clinicians must use the test if they want to have the respective therapy reimbursed.

Apart from the tests for these new targeted therapies, on which I will dwell upon later in this chapter, as of today, a clinician in charge of making such treatment decisions, namely the oncologists, do not have much more at hand as to base these decisions mainly on the age and any health status of the patient as well as on the type and extent of the tumor (staging). Staging takes into account the size of a tumor, how deep it has penetrated, whether it has invaded adjacent organs, if and how many lymph nodes it has metastasized to, and whether it has spread to distant organs. According to these parameters, a cancer is classified into Stages that range from I to IV, with increasing size and aggressiveness. In this respect, the staging of cancer is currently the largest predictor of survival (prognosis), and directly influences treatment decisions. Early stage cancers (Stages I to III) are usually localized to the organ of origin, and patients may receive surgery or radiation as the primary treatment. After this primary treatment, the patients may be offered chemotherapy in order to reduce the chances that the cancer will return ("recurrence"). This therapy, given after primary treatment, is called "adjuvant therapy." In later stages, where the cancer has spread to other organs or when the cancer has recurred, patients are given a so-called palliative treatment. Here, the expectation is

rather to reduce the severity of cancer symptoms, improving the quality of life for the patients and their families and, whenever possible, to prolong life.

Thus, staging sets the frame regarding what group of treatment alternatives an oncologist could consider, but it reveals nothing about which of the treatment alternatives might be beneficial for the patient and which is not. Results from clinical trials, along with experience, surely provide an oncologist with overall benefit and side-effect profiles; although it does not allow them to make predictions for a specific individual. This results in the situation that oncologists have to put all their patients at risk of experiencing side effects with quite often only a minority of the patients benefiting from the treatment.

1.3 Pharmacodiagnostic Testing

A proper solution to this problem could come from the upcoming field of pharmacodiagnostic testing. Pharmacodiagnostic is a fairly old discipline, although most of its possible implications are still in the developmental phase, with antecedents that stretch as far back as to the beginning of the twentieth century (Goldstein, Tate, & Sisodiya, 2003). Pharmacodiagnostic studies the genomic basis of interindividual variability in the response to drug therapy, and, in this way, seeks to develop more rational means for optimizing drug-related therapy by ensuring maximum efficacy with minimal side effects with respect to the patients' genotype. In the last five decades, due to numerous studies, it has become apparent that each drug interacts with variant genes and proteins (transporter, enzymes, binding proteins, and receptors) in the body, and, moreover, affects hundreds of proteins in metabolic and signaling pathways downstream of the primary interaction (Sadée & Dai, 2002). Approaches, such as those of pharmacodiagnostic, promise the advent of "personalized medicine" in which drugs and drug combinations are optimized for each individual's unique genetic composition. This approach was especially promising for those therapies that are associated with severe side effects, such as most cancer therapies. The appearance and behavior under certain conditions of healthy as well as abnormal cells, such as cancer cells, are determined by human beings' genome (Jones & Laird, 1999), which makes it reasonable to apply the approach of pharmacodiagnostic to the general field of oncology. Since terms, such as pharmacodiagnostic, pharmacogenomic, pharmacogenetic, and a good deal more, tend to be used interchangeably, and a precise consensus definition of either remains elusive, for this thesis, pharmacodiagnostic encompasses every kind of testing that deals with the influence of

any, either genetic or genomic, variation on drug response and eligibility as well as the severity of the disease or drug toxicity.

Tab. 1: Different Test Models in the Field of Pharmacodiagnostic

<i>Test</i>	<i>Focus on what?</i>	<i>Description</i>
<i>Responder test</i>	Drug response	Test result notifies whether treatment X benefits the patient.
<i>Target test</i>	Treatment selection	Test result notifies whether the target for the drug is present in the tumor, which is an indication for potential response, but not necessarily correlated.
<i>Adverse drug reaction test</i>	Adverse drug reaction (side effects)	Test result notifies whether the patient has a higher risk of suffering from adverse drug reactions.
<i>Prognostic test</i>	Severity of the disease	Test result notifies whether the patient has a good or bad prognosis, independent of the treatment. However, it is used as an indicator for how aggressive the treatment needs to be.

The range of possible effects an individual's genotype—that is, the genetic composition of each person—can have on drug-related issues has suggested different market models of tests. These models can be roughly distinguished in responder tests, target tests, adverse drug reaction tests, and prognostic tests, all of which are described in more detail in Table 1.1. The most prominent examples of pharmacodiagnostic testing in the field of oncology that are already routinely applied in current clinical practice are certainly several target tests coming with the respective targeted therapies. The results of these tests should notify the clinician whether or not a patient might benefit from the respective targeted drug, as the gene target for the drug is expressed in the tumor. A very well-known example of such a target test is the *Herceptest*[®] for the drug Herceptin (*Trastuzumab*[®]), which has been approved for a subpopulation of advanced breast cancer patients that overexpress the *HER-2* receptor (Goldstein et al., 2003). Beyond this and some other examples, there are not so many other tests in the aforementioned categories. One of the newest examples in the field is the

OncotypDx[®] test, a prognostic test for breast cancer patients, applied after primary treatment with surgery in order to better tailor further administration of treatment. However, a breakthrough to clinical routine is still awaited for this test.

Without a doubt, pharmacodiagnostic tests have the potential to more individualize the administration of already existing therapies with the clear intent of maximizing effectiveness for both patients' health and survival as well as for the health system by saving costs for expensive medicines that have no beneficiary effects for many patients or that are associated with adverse drug reactions. Moreover, pharmacodiagnostic can hold a promise to the pharmaceutical industry by offering useful tools to improve tremendously expensive drug discovery and development processes by better selecting the patients who are eligible for clinical trials.

Thus, the benefits of pharmacodiagnostic appear to be “just around the corner,” while “costs” seem to be deemed a distant project. Indeed, the general literature on pharmacodiagnostic is characterized by a lack of critical edge. Therefore, Holtzman (2003) appears an almost lonely figure when, also for the field of pharmacodiagnostic, he points at issues, such as poor predictive value performances of celebrated test examples, in the literature. By now, it is not much recognized in the respective literature that, in fact, pharmacodiagnostic tests are tainted with the same drawbacks as any other diagnostic tests, that is, the delivery of incorrect results. When ordering a low-performing test or misinterpreting test results during decision making, pharmacodiagnostic tests can be more misleading than informative. That is, before utilizing a test, each clinician has to consider whether the information provided by the test is at all helpful. That this notion is by no means ill-founded is proven by several studies showing the difficulties clinicians have with respect to a correct interpretation of diagnostic test results. For instance, when clinicians were asked to draw inferences from the positive predictive value (PPV) of tests, they were incorrect in many cases (Casscells, Schoenberger, & Grayboys, 1978; Hoffrage & Gigerenzer, 1996) or were confused with the sensitivity of a test, and, in addition, rated false alarms of tests as being quite rare or even nonexistent in cases of repeated testing (Gigerenzer, Hoffrage, & Ebert, 1998). For instance, in the case of a prognostic test in breast cancer, the emphasis of the result of such a pharmacodiagnostic test might not even be so much on false positive test results³. In

³ Given the result, it is falsely believed that the patient has an aggressive tumor, and, therefore, is treated with more toxic treatment options.

this case, it would simply reflect the kind of overtreatment already ongoing in medicine. The patient would be offered treatment as defined currently in the guidelines although it is recognized that, in many cases, patients are overtreated and should be better off with less toxic treatment options. The risk of having a false negative result⁴ is much more problematic. Not to administer a potentially beneficial therapy to a patient who is in urgent need, due to a false positive result, could have a dramatic impact on a patient's recovery, quality of life, or, even worse, on their survival.

1.4 The Goal of the Thesis

The field of pharmacodiagnostic is on the verge of expanding rapidly, and soon an abundance of various tests might hit the market and guide the oncologists' treatment decision making by notifying them about patients' drug response and eligibility, their risk of experiencing severe side effects, or whether or not they are at bad prognosis. It was highlighted that pharmacodiagnostic tests could hold great promise for future medicine by offering the foundation for more individualized therapies for patients, and for better tailored drug discovery and development. However, there are also pitfalls and drawbacks, and, up to now, they are not stressed in the literature as the potential benefits.

To my knowledge, no study has addressed the issue of how clinicians would handle this new technology of pharmacodiagnostic tests within their daily practical routine, and to what extent an awareness of the pitfalls of that specific technology exists amongst them. This is astonishing, given the potential that negative impact pharmacodiagnostic tests can have on patients' lives, and, thus, also on the health system. However, obtaining sufficient insight into this topic could elicit problems that might occur with the application of such tests at an early stage of their introduction to clinical practice. According to this, such studies may serve as a reasonable basis for the development of adequate training tools and decision aids for medical care personnel likely to apply such pharmacodiagnostic tests. Since most of these tests focus on therapies administered for severe diseases, such as cancer, the field of oncology seems to offer a good representative field to commence and conduct an elaborative study to acquire more knowledge about the application conditions of pharmacodiagnostic tests. In the field of cancer, decision making is fundamental to all aspects of care, yet researchers and clinicians

⁴ Given the result, it is falsely believed that the patient has a less aggressive tumor, and, therefore, they need only to be treated with a less toxic treatment.

have limited knowledge of the way in which patients and their health providers make these critical health decisions (Nelson, Stefanek, Peter, & McCaul, 2005).

Thus, this thesis is a first effort to examine how oncologists would deal with such a test when facing a decision on it in their daily practice. In a first and explorative step, it is aimed at finding out whether the consideration of potential risks and benefits of such tests or rather something else would drive a decision on applying such a test. That is, the first step is dedicated to the elicitation of cues that are important for such a decision. In a second step, it is explored, with methods of judgment analysis, how these cues are used in order to come to a final decision of applying such tests in clinical routines. In accordance with the idea of shared decision making for any medical decisions, it also is examined what role patients play in the decision-making process on these tests, and what cues would drive their decision on wanting to see such tests applied to their treatment decisions. Furthermore, it also is investigate, for patients with methods of judgment analysis, how these cues would be used in order to come to a final decision.

By taking decision making as the general approach of this thesis, the subsequent aim in applying methods of judgment analysis is to elicit whether the nature of the decision-making strategies of the respective groups are better explained by a less complex and noncompensatory⁵ decision-making model than by a more complex and compensatory mechanism that combines and weights different information.

Finally, it was speculated whether nationality could have an impact on choice. A good comparator seemed to be the USA. The USA is often regarded as being scientifically years ahead, compared to Germany, which might make a huge difference when focusing on an innovative technology, such as pharmacodiagnostic tests, and is furthermore, in the sector of medicine, seen as being more aggressive in applying therapies that might hold even the slightest promise of being beneficial for the patient, which is partly explainable by liability issues. That is, the third goal of this PhD thesis will be to investigate if there are any significant group differences regarding the choice behavior with respect to groups' nationalities.

⁵ A decision strategy is noncompensatory if the value of one cue cannot be outweighed by the value(s) of one or more other cues.

2 When Physicians Have to Decide: A Case of Complexity?

“To those who are medically naive, the
complexity of modern medicine can sometimes
appear overwhelming.”

Steven Schwartz & Timothy Griffin

Making and prescribing decisions are a physician's major activity. Every day, all over the world, patients describe their symptoms and complaints to their doctors while these perform examinations, order tests and evaluate their results in order to make a decision, on the basis of all this data, on how to treat the respective patient. The final decision is preceded by numerous very carefully taken and reasonable prior decisions, which enable doctors to come up finally with the solution for the overall question on the most appropriate treatment, out of a set of several, including the alternative of no treatment. This is how it appears to all of us and how we want it to appear. Our doctors make decisions on patients' lives and well-being every single day, which sometimes includes ours, and that is what makes us want to believe that apart from a nearly infinite knowledge about every medical fact, the decision making of our doctors includes the ability of skilled and elaborated clinical judgment. This might be less astonishing, as in many medical situations the decisions made by our doctors are usually sold to us as being a very defined and safe decision, which surely has its roots partly in the patient's desire for certainty. But is this imagined medical certainty really so certain? Furthermore, what does it mean to this certainty when one or more test results have to be considered, and to be involved in that process of forming a final medical decision?

Various studies in medicine have found wide variation among physicians in their judgment (e.g. Strauss, Chassin, & Lock, 1995). Both inferred cue weights as well as applied strategies often vary considerably from each other or those recommended in the literature. For example, in a study of the diagnosis of pulmonary embolism conducted by Wigton, Hoellereich, & Patil (1986), it was found that the relative weight given by experienced internal medicine faculty members to the level of oxygen in the blood varied from no weight at all to 90%, although this cue was strongly emphasized by the textbooks of that time. In another example, Kirwan, Chaput de Saintonge, Joyce, & Currey (1984) found greatly varying estimates when asking rheumatologists to estimate the change in disease activity in a series of

cases with patients suffering from rheumatoid arthritis. Quite comparable findings were reported by Poses et al. (1992) when physicians were confronted with a task on diagnosing strep throat of patients with pharyngitis. Such variations in judgment are not only exclusive to diagnosis but also to judgments about therapies. Agreement between 70 psychiatrists in selecting which drug to prescribe for hypothetical psychiatric patients was no better than chance level (Gillis, Lipkin, & Moran, 1981), and Evans, Harries, Dennis, and Dean (1995) showed that physicians' policies were highly variable in using risk factors and clinical information to select a drug for hyperlipidemia. Seeking an explanation, it was also thought that it might be a function of being either expert or novice, with the underlying idea that, with increasing experience, experts' cue weights should become more alike. However, the hope for obtaining at least one good explanation for those variations in clinicians' judgment did not prove true (Slovic, Rorer, & Hoffman, 1971; Wigton, et al., 1986; Poses, Bekes, Winkler, Scott, & Copare, 1990).

Such findings seem to be at odds with the common assumption of our doctor's judgmental ability, and, at first glance, are not very tempting. But before deciding never to go to a doctor again due to the uncertainty apparently inherent in their decisions, let us find out a bit more about clinicians' decision making. By doing so, we might discover on the stormy sea that we have done pretty well with our doctors during the last decade, without much certainty.

2.1 Some Words About Certain Uncertainty in Medicine

Without a doubt, today's doctor's working life is complex and probabilistic. They have to access a large variety of clinical data every day. These data, gathered from physical examination as well as from various tests, are considered to provide them with information about the true health condition of their patients. However, since these tools are just for measuring a particular indicator of the health condition and not the health condition itself, information derived from this is not a 100% correct copy of the reality, but simply probabilistic information. None of the available examination procedures or any of the available tests is able to discriminate perfectly between a truly ill and a healthy patient. In addition, symptoms' and signs' relevance for a specific disease and their relationship to it are not known for quite a majority of diagnostic and treatment judgments in the field of medicine. For instance, the signs and symptoms of a heart attack can differ in relation to the age of the

patient, their ethnicity, and the nature of the injury to the heart (Wigton, 1996). This points towards the fact that medical decisions will always be less than perfect.

Over decades, a huge amount of medicine's research effort has been guided at providing increasingly more information. The sheer masses of diagnostic tools available nowadays are witness to the success of this effort. However, while one might hope to clear the sky of uncertainty by having more and more diagnostic tools, such as tests at hand, these tools may partly be the source of new clouds in a sky that was never clear. Since these tools are now available, doctors are expected to make use of them regardless of whether they are able to handle and interpret these appropriately. Clinicians seem to have no choice; they must somehow learn to deal with an enormous potential and probabilistic database to make management of care decisions for their patients.

However, most human beings, including doctors, have great difficulty in thinking probabilistically. The mathematical complexity of probabilistic calculations and, additionally, in the case of medical decisions, the amount of information that must be evaluated seems to lead into a jungle where is no way out. By now, medical training and textbooks might not be of much help for doctors in finding a sound solution for how to better deal with probabilistic test information.

2.2 Condition of Medical Decision-Making Tasks

2.2.1 Decisions on Tests: A case of Task Complexity or the Bite of Bayes

When going through the medical literature charged with advising medical trainees, the questions of how to decide when to test and how to understand and interpret a test's results usually pose quite challenging ideas. The probably most well-known and most often taught formula for deciding when to test and for interpreting test results in the course of medical training is the normative approach⁶ of *Bayes' theorem*. This theorem is a description for logical reasoning in order to find the mathematical optimum when evaluating hypotheses. It states that the conditional probability that a patient suffers from a disease given a positive test result is a function of the prevalence of a disease, the conditional probability of a positive test result given that the patient actually has the disease, and the prevalence and conditional

⁶ The definition of 'normative' refers to theory, which describes how a perfectly rational, super-being *would act*. In contrast, a descriptive theory does describe how a person *acts*, while a prescriptive theory prescribes how a person *should act*, what can be close to the normative ideas but do not have to. (Freeling, 1984)

probabilities associated with the alternative hypothesis, that is, no disease while still showing a positive test result. In the case of deciding whether or not to test, each physician would have to ask themselves whether the received test result would change the management of care considered for a patient, or, to phrase it in another way, whether the test can reduce uncertainty. Let us introduce Dr. Plaster to make the following ideas more clear. Dr. Plaster is faced with a patient he suspects to have a tropical disease. To decide on whether he should conduct an available screening test for that disease, Dr. Plaster first has to calculate a treatment threshold probability, that is, the probability of disease where Dr. Plaster would be indifferent between giving treatment and withholding treatment. Several methods are recommended in the medical literature for doing this. Apart from the opportunity of determining intuitively what benefits and harms of giving the considered treatment could entail for the patient, given that they have the disease or not, Dr. Plaster could further conduct fancy and very long mathematical computations for calculating cost/benefit ratios which can include either life expectancy or a patient's utility estimates. As to us, it does not appear reasonable at all to assume that any practical working doctor has the time for making such long calculations, and remember Dr. Plaster would have to do this for every single patient, therefore, I am not going to dwell on these equations here,⁷ but rather assume that Dr. Plaster decides to come up with an intuitively completed probability estimation for the treatment threshold. In the case of his patient, he decided that he would treat with, for example, penicillin, if the probability of having the tropical disease was at least .20. This means that Dr. Plaster would treat if the probability of the disease was above .20 and withhold treatment if the probability was below .20. Let us return to our tropical disease, Example 2, and that Dr. Plaster assumes that the probability of the patient having the disease is .30, which is again made more or less intuitively. At this point, Dr. Plaster would treat his patient because the disease probability is above the set treatment threshold. But he has now to consider if a test result could affect his decision to treat. To answer this question, Dr. Plaster must decide if a negative test result (indicating that the patient does not have the tropical disease) could reduce the upfront intuitively estimated disease probability (.30) to less than .20 (treatment threshold); otherwise he would not change the already considered management of care. This would be the moment when our smart Dr. Plaster would introduce Bayes' theorem to his calculations. To

⁷ For interested readers, I can recommend the book from Sox, Blatt, Higgins, & Marton (1988) "Medical Decision Making" where it is exhaustively outlined how to calculate treatment threshold probabilities.

work with this theorem, Dr. Plaster needs to know a little more about the screening test's performance in terms of possible outcomes. In general, every test can have four possible outcomes. When the patient has the disease (*disease+*), the test can be either positive, which would be correct and therefore is called *true positive* outcome, or the test could be negative, which would be incorrect, and is therefore called *false negative* outcome. These first two outcomes are linked to the *sensitivity* of a test, the measure of the proportion of patients who test positive given the disease. For instance, out of 100 patients, 10 patients do have the tropical disease (*disease+*) and test's sensitivity is 80%. When focusing on these 10 patients having the disease, 8 persons will have a positive test result (*true positive*) and two will have a negative result (*false negative*). The rates of these two outcomes always total 100%. When the patient does not have the disease (*disease-*), the test can again be either positive, which would be incorrect in this case and is therefore called a *false positive* outcome, or the test could be correctly negative, which therefore is called *true negative* outcome. These two outcomes are linked to the *specificity* of a test, the measure of the proportion of patients who test negative given the absence of the disease. To provide another example: It was mentioned above that, out of 100 patients, 10 do have the tropical disease, which means that 90 patients do not (*disease-*). If the test's specificity was also 80%, then from these 90 patients, 72 persons will have a correct negative test result (*true negative*), while 18 will have an incorrect positive result (*false positive*). As it can be seen here again, these two outcomes total 100%. Since Dr. Plaster is always up-to-date and reads the latest well-appreciated medical journals, he knows that, for the tropical disease screening test, both test's sensitivity as well as test's specificity turned out to be 80% in clinical trials. Equipped with this, he has now everything he needs to calculate the posttest probability of a patient given a negative test result. Only when the posttest probability of the disease would be below the threshold it would make sense to apply the test as the result would alter the management of care to not treating the patient. For determining how likely the patient still is to have the disease (posttest probability) given a negative test result, Dr. Plaster must add in the Bayes theorem formula (Formula 1) some of the values mentioned before. The respective values he needed for this equation were the pretest probability of the disease, and, in addition, the probability of the test incorrectly negative when the disease is present (*false negative*) as well as the probability that the test is correctly negative when the disease is absent (*true negative*) (see Formula 2).

2. When Physicians Have to Decide: A Case of Complexity?

$$P(\text{Tropical D/Test-}) = \frac{P(\text{Tropical D}) \times P(\text{Test-/Tropical D})}{P(\text{Tropical D}) \times P(\text{Test-/Tropical D}) + P(\text{Not Tropical D}) \times P(\text{Test-/Not Tropical D})} \quad (1)$$

$$P(\text{Tropical disease/negative test}) = \frac{P(.30) \times P(.20)}{P(.30) \times P(.20) + P(.70) \times P(.80)} = .10 \quad (2)$$

At last, Dr. Plaster had found out that it would be worth doing the test, as the negative test result moves the posttest probability of the disease with .10 (2) under the treatment threshold and therefore would change Dr. Plaster's method of treating the patient. That is how the theory works. The enormous amount of money spent every year for diagnostic testing,⁸ of which a certain amount of this is assumed to be carried out totally unnecessarily, might be witness that this calculation has its problems. It is apparent that Bayes' formula follows the laws of logical mathematics and rationality, but clinicians clearly do not revise probabilities in line with the requirements of Bayes' theorem. In fact, there is empirical evidence that doctors are widely unable to understand and work with the Bayesian formula. For instance, confusion about probabilities has been explored by Eddy (1982), whose research focused on doctors' interpretation of mammography test results. Eddy used a hypothetical, but common, situation as his experimental task, and asked physicians to estimate the probability that a woman has breast cancer given a positive mammogram. He provided his participants with all the probabilistic information, as were used in the example above. Out of a sample of 100 physicians, 95 estimated the probability 10 times higher than it actually was. This estimate appeared to be based solely on information provided by the test's true-positive rate (sensitivity), while ignoring both pretest probability of the disease and false-positives. Similar findings have been reported by Schwartz, Gorry, Kassirer, and Essig (1973) as well as by Gigerenzer et al. (1998). As the Gigerenzer and Hoffrage (1995) stated, even if scientifically working clinicians *would* understand Bayesian formula, ordinary people *and average physicians* rarely do (*italic is added by the author*). Might more training for clinicians in

⁸ The amount of money spent for performing laboratory testing increased in Germany from €4.30 million in 1992 to €5.93 million in 2004 (Statistisches Bundesamt Deutschland, 2006).

Bayes' rule be a solution? One might lose the game betting on that horse. A study from Sedlmeier and Gigerenzer (2001) found that 1 week after students had successfully passed such a course, their performance was already down by 50%, and it continued to fade week after week. Moreover, they stated that the chance of convincing a physician to take a statistics course in the first place was almost nil, most had no time or little motivation, while others believed they were incurably innumerate.

We have to keep in mind that deciding on when to test and interpreting tests' results are far from being all the steps needed to be taken for a final diagnostic or treatment decision. There is empirical evidence that when decisions become more complex, people will tend to use simple decision strategies. A series of experiments indicate that choice strategies are sensitive to the number of alternatives between which to choose (e.g., Shields, 1980; Klayman, 1985; Olshavsky, 1979; Onken, Hastie, & Revelle, 1985; Payne, & Braunstein, 1978; Payne, Bettman, & Johnson, 1993). When people are faced with multialternative decisions, they seem to respond to such complex tasks by applying simple strategies where they just focus on one fact (noncompensatory strategy) to make the decision (e.g., Biggs, Bedard, Gaber, & Linsmeier, 1985; Sundstrom, 1987; Chinburapa et al., 1993).

2.2.2 *Do Pharmacodiagnostic Tests Have the Potential to Smart Task Complexity and Uncertainty?*

The pharmacodiagnostic tests' underlying main idea certainly is to give physicians a more objective decision aid for more precise treatment decisions at hand. This seems to be especially favorable for such diseases with low survival rates and treatments entailing severe bodily harm for patients. However, it is reasonable to assume that already mentioned findings suggesting serious problems in physicians' understanding and interpreting of probabilistic test information are as true for pharmacodiagnostic tests as they are for any other given diagnostic test. Furthermore, it is assumed that also for such tests it is difficult for clinicians to correctly calculate when to test would be appropriate and when it would not. That is, pharmacodiagnostic tests will certainly have their benefits, but surely will not smart the task complexity or uncertainty of physicians daily work practice. Beyond the task complexity that physicians commonly face, they furthermore have to deal with another issue, namely, time pressure.

2.2.3 *Time Pressure*

Physicians are often faced with an intense workload, and, additionally, are rushed not to spend more than only a few minutes⁹ with one single patient due to tightfisted reimbursement conditions, which certainly lead to an implicit, if not even explicit, feeling of time pressure. That is, today's Dr. Plaster is asked to make decisions rapidly and will almost surely lack the time necessary to make calculations in line with the Bayes' theorem when having to decide on testing. Research has revealed that time pressure affects people's judgment and decision making in many ways. It has been demonstrated that humans do not use all relevant information (e.g., Rothstein, 1986; Wright, 1974). Wallsten and Barton (1982) as well as Payne, Bettman, James & Johnson (1988) reported that the focus of attention shifts to more important information under the condition of time pressure. While Edland's (1989, 1994) findings confirmed that the effect of this time pressure resulted in greater selectivity of information use, moreover, she found greater use of noncompensatory strategies. Rieskamp and Hoffrage's (1999) study results strengthen that – under time pressure conditions – individuals' judgment seemed to be better described by a simple noncompensatory strategy, which was characterized by little information search and a decision based on a single cue.

2.2.4 *Quality of Information*

Medical information is gathered by examination and testing, and is probabilistic and noisy. When Dr. Plaster wants to find out about the patient's true state, he has to accept that it usually cannot be directly observed (Sox, Blatt, Higgins, & Marton, 1988), and so he must use external imperfect cues, for example, test results, and try to infer the patient's true state. Indeed, clinicians cannot be certain what a finding implies about the patient's true state. But even when information is available to Dr. Plaster, he quite often does not exactly know how useful this information is in predicting whether a patient actually has the disease or not, as weights are often difficult to obtain due to their situation-dependence. Although during their medical training, physicians are taught which symptoms represent which kind of disease, they usually are not provided with information about a ranked order of the predictive validity of each of the symptoms for a respective disease (Wigton, 1996). This is mainly due to the fact of the occurrence, the type and the strength of the symptoms is dependent on several conditions

⁹ On average, a German general practitioner sees his/her patient less than 6 minutes per encounter (Engelhardt, 2006)

of the patient, such as age, gender, or co-morbidities. In fact, for most situations in medicine, it is almost impossible to measure exactly the predictive validity of symptoms regarding the outcome.

2.2.5 *Outcome Feedback*

A way of establishing at least the subjective predictive validity of different pieces of information would be for a clinician to gather feedback about the outcome of their decisions. Obtaining the outcome after a medical action is taken based on a diagnosis seems to be easy to achieve. This feedback can improve the achievement of medical judgment, and is empirically well-proven by different cognitive feedback studies (e.g., Hammond, 1971; Wigton, 1987; Poses et al., 1992). However, in these studies, physicians are presented with correct symptom/actual state relationships, while, in reality, the obstacle of making use of outcome feedback exactly comes up with the rather often unknown relationship between symptoms and the actual state of a patient. For instance, if one of Dr. Plaster's patients had a bladder inflammation, but Dr. Plaster was to diagnose it as a nervous stomach, since bladder infection and a nervous stomach shares some identical symptoms, therefore, he would treat it as such. Over the course of a few days, the bladder infection disappears by itself. The next time Dr. Plaster meets this patient, he would be informed how well the treatment worked, and Dr. Plaster will memorize that the information leading him to the final diagnosis was useful for a nervous stomach. That is, physicians rather often do not know if they used the right information in the right way. It is apparent that learning from unfounded experience might not be the best source for drawing inferences, as it is prone to bias and error (Brehmer, 1980).

2.2.6 *Cost of Health Care*

A present day medical decision is truly not only driven by physicians' free will or rational decision process. Health care is expensive, and the amount spent on this every year continues to rise fast. There are several reasons usually offered for the high costs of health care: new expensive technology, the aging population who require more medical attention than previous generations, doctors' increasing fees, and a lot more than the absolute necessary ordering of tests and other procedures due to an increasing fear of lawsuits (see Gay & Sax Jacobs, 1983), especially in the USA. Furthermore, even if it is usually not listed as a skyrocketing reason for escalating medical costs, it is quite obvious that public and

professional attitudes toward health and illness are surely among the most important causes for the current situation (Schwartz & Griffin, 1986). The high cost of health care definitely mirrors widely held ideas of what constitutes desirable medical practices, sometimes regardless of how small the increment may be (e.g., routine mammography screening, Gigerenzer, 2002).

However, there are more and more limitations set by health insurances as escalating costs are becoming a problem for societies in the Western world. Physicians can no longer administer every available treatment and diagnostic tool to the patient in a way they would deem the best. Which one they could, and which one they could not, is increasingly a matter of present budgets or reimbursements, both of which are clearly defined by the respective health insurance of the patient. For instance, in Germany the quite cost-intensive drug *Herceptin* is only administered for governmental health insured breast cancer patients after being diagnosed as metastatic, while privately insured breast cancer patients receive this drug already in the earlier and better curable nonmetastatic stage¹⁰. To have the treatment reimbursed, physicians have to perform the related target test. That is, even if Dr. Plaster had his concerns in applying this test, he would not have had much choice whether to perform it or not, as without the positive test result the respective treatment would simply not be reimbursed. Although between Germany and the USA there are differences in the health system, that is, while German health insurances tend to provide more and more defined budgets for single diseases, US health insurances usually reimburse defined single medical actions, the general question and the inherent limitation of who is paying the bills for each medical decision remains the same in both countries.

2.2.7 Medical Guidelines

The introduction of clinical guidelines to the field of medicine¹¹ has been seen as a way out of the complexity, variance, and malpractice. While in Germany it is still quite a new decision aid for several specific diseases, and in the field of oncology is still rather in its infancy, in the USA guidelines are already well-established. As the name already indicates,

¹⁰ In both cases, the respective test mentioned in Chapter I has to be performed, and only a positive result yields administration of the drug.

¹¹ In Germany, the *German Society for Cancer* (DKG-ISTO) in cooperation with the Working Committee of scientific medicine (AWMF) has been in charge of developing and publishing guidelines in the field of oncology since 1995.

guidelines are supposed to help physicians in guiding them toward proper diagnosing and making treatment decisions by offering them logical rules of how to proceed step by step under several conditions. Guidelines are assumed to pose the current gold standard in diagnosing and/or treating diseases, respectively, and everything that is suggested in their claims have to be proven as being advantageous, compared to other procedures, by a certain degree of empirical evidence¹².

However, doctors seem to have difficulties to come to terms with following strict rules. Ustun and Sartorius (1995) reported underdiagnosis of depression despite the dissemination of many clinical guidelines to guide recognition and treatment of the disorder. Smith, Gilhooly, & Walker (2003) findings in their study conducted with 62 general practitioners indicated that the doctors did not seem to be following the logical rules embodied in the guidelines, which they assumed might have been an issue of its complexity.

2.2.8 *Last But Not Least: Patients' Meddling In*

When nowadays Dr. Plaster is about to make a medical decision for one of his patients, he is expected to do this slightly differently than he would have done back in the 1980s. Although still a sound amount of people tend to regard their doctors as being in charge of making proper medical decisions solely, attitudes about what good medical decision making means have shifted in recent years away from the ideal of the doctor's paternalistic responsibility toward shared decision making. Today, in a doctor-patient encounter, shared decision making is regarded as the standard, and doctors are more and more urged to ensure that this occurs in their practice (Butow & Tattersall, 2005). That is, whatever Dr. Plaster might have conceived for his patient as being the best, is increasingly likely not to come to fruition, since the patient might have preferences different from the doctor's proposed type of care. For potentially life-threatening diseases, such as cancer, which often involve weighing up uncertain benefits against uncertain side effects, there is some evidence that most patients still want their physicians to take the decision for them. I am going to dwell on this topic more closely in Chapter V.

¹² That this holds not necessarily true is documented by a case in which the respective test for the drug *Herceptin* (see also Chapter I) received approval by the Food and Drug Administration (FDA) based on the pressure of breast cancer patient groups, who hoped to accelerate the launch of the promising drug.

2.2.9 *Unbounded Bounded?*

It is apparent that medical decisions are based on the grounds of complexity and uncertainty. Dr. Plaster's working life and those of his colleagues seems to be best described by heavy workload, the unknown patient's true state, and fog around probabilistic test information as well as around the predictive validity of a single piece of information. Conditions perfectly made for directing physicians into medical inferences are prone to bias, inconsistency, and wide variation, amongst them. Many assumptions and models have been proposed over the last decades from many disciplines, describing how humans deal (descriptive models), and should deal (prescriptive and normative models), with the abundance of information in order to make an optimal choice. Most of these prescriptive and normative approaches have been shaped by the appealing idea of *unbounded rationality* that expects human beings to be stuffed with perfect knowledge, gigantic computational power, and infinite time for making all the computations needed for processing and weighing all gathered information. Unbounded rationality is traditionally modeled by the probability theory, which did not become institutionalized until the 1950s. Its best known realizations are surely the *maximization of expected utility* by Bernoulli and *Bayesian models* (Gigerenzer & Todd, 1999). Imagine Dr. Plaster again attempting to resolve the question of whether to apply a pharmacodiagnostic test. To calculate his, or in this case probably more the patient's, expected utility for testing, he would have to determine all the possible consequences that testing could bring (e.g., no overtreatment, no side effects, the risk of undertreatment, the risk of wasting money for the test), attach quantitative probabilities to each of the consequences, estimate the subjective utility by its associated probability, and, finally, total these numbers. He would have to carry out the same procedure for the other alternative, namely, "not to test". Eventually, he would end up with two estimates, and would have to choose the alternative with the higher total expected value. It is apparent that Dr. Plaster, in order to obtain all the information about the consequences, their probabilities, as well as their utilities, might have to invest years in coming up with one accurate test decision. It would be interesting to see which hospital would be willing to pay for such an accurate, but quite inefficient, working physician.

Because of the unnaturalness and ignorance of constraints imposed on human beings, unbounded rationality has come under criticism. Considerations of limited cognition, such as memory span and knowledge, together with the subsequent costs of required search for information, form the main approach of *optimization under constraints*. The major difference

between the maximization of expected utility and optimization under constraints certainly is the appreciated limited knowledge as well as the limited information search for the latter model. Although in optimization under constraints search is limited and perfect knowledge is dropped, it retains the notion of optimization, and that is where Dr. Plaster gets into trouble again. Suppose that Dr. Plaster, who attempts to optimize under constraints, has already listed two consequences of pharmacodiagnostic testing, for example, the potential of avoiding overtreatment and of avoiding side effects from the treatment, as well as estimating their respective probabilities and utilities. Before he proceeds to a third consequence, he must calculate whether the benefits of continuing this information search will outweigh its costs. If not, Dr. Plaster could stop his search immediately. However, to compute the benefits of further search, the tortured Dr. Plaster would have to reconsider what all the third consequences could be, again estimate their utilities and probabilities, compute how much each could change his eventual decision, and average all of these to finally come up with the expected benefit of continuing search. But Dr. Plaster, who tries to optimize under constraints, is not finished yet, as he would also have to calculate the costs of continuing the search as. This would also include such considerations as what he could have been doing during the time he used considering his decision. As one can see, in the end it leads to a never-ending procedure, and again no hospital would be prepared to employ the optimized Dr. Plaster.

Although due to its original assumption of limited knowledge and limited search, optimization under constraints has often been mistaken as belonging to models of *bounded rationality*, but actually has also to be categorized under those unbounded ones. The models encompassed under the conception of unbounded rationality propose decision strategies, best described as being compensatory, since cue values can be outweighed by other ones (Rieskamp & Hoffrage, 1999), and as search-unlimited, since no (useful) rule of when to best stop search is suggested. In this, the concept of unbounded rationality has little or no regard for constraints of knowledge, time, and computational capacities that physicians usually face in daily life. They often have no systematic way of using all the gathered medical information, they do not have a reasonable way in searching rationally and logically for new ones (see Wertman, Sostrin, Pavlova, & Lundberg, 1980), and so find themselves relying again on intuition and ad hoc decision rules, being distant from behaving optimally or even rationally as Bayes' theorem would suggest.

2.2.10 *Clearing Up the Sky by Offering Bounds*

A great deal of psychological research has been to understand and clarify how we actually make our quite accurate decisions and judgments under the constraints of our limited knowledge, processing capacities, and also usually under limited time. In recent decades, psychologists have come to believe that we fail to conform not only to Bayes' theorem but also to all of the unbounded rational models because we simply lack the cognitive capacity to *search and* combine the required information in a systematic manner (Schwartz & Griffin, 1986) (*italic is added by author*). Instead, the idea of *bounded rationality* was proposed. The father of bounded rationality surely is Herbert Simon. In his book *Models of Man: Social and Rational* (Simon, 1957) he stated: "The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solution is required for objectively rational behavior in the real world or even for a reasonable approximation to such objective rationality." (p.18) Simon (1956) argued that normative models of decision making ignored situational and individual constraints, such as time and information processing capacity. Simon proposed that people develop shortcut strategies, such as simplifying *heuristics*, which deliver reasonable solutions to real-world problems, and that humans are rational within the bounds of limited information-processing capacity, that is, they are bounded rational. This picture of limited human judgment ability may have not appeared very flattering to everybody. It therefore might be not surprising that bounded rationality's emphasis of humans, cognitively and timely limited, has been ever since quite often misphrased to humans being flawed and irrational in their behavior and decisions.

2.2.11 *Heuristic = Bias?*

A program that partly grew from the view of Simon has been the *heuristic and biases* program of Kahneman and Tversky (1982). Although the idea of this program disagrees with rational theories on whether humans follow norms of rationality, it still assumes that we are basically rational decision makers and, therefore, does not question the norm of bounded rationality itself. In the authors view, humans tend to quite often apply heuristics in inappropriate situations, therefore, leading to errors to which the authors refer to as cognitive "biases." The major three proposed heuristics that are also assimilated from the field of medical judgment investigators are *representativeness*, *availability*, as well as *anchoring and*

*adjustment*¹³. Their program has produced a long list of biases and has also shaped many fields as well as medical judgment. One of the strategies to defeat the burdens of heuristics has been to recommend debiasing methods (e.g., Plous, 1993; Hammond, Keeney, & Raiffa, 1998) for overcoming biases, such as representativeness. However, while debiasing methods have been demonstrated in the laboratory (e.g., Arkes, Christensen, Lai, & Blumer, 1987; Lichtenstein & Fischhoff, 1980), the methods have not been shown to remarkably improve decision making in the field setting, where it actually would be useful. Data suggest that decision biases do not even have a noteworthy influence in the field setting (Fraser, Smith, & Smith, 1992). Moreover, the strategy of debiasing is contradictory to the concept of heuristics, as it forces people to leave the way of being “simple” and plunge them into demanding computational processing again. The promotion of debiasing illustrates once more the impact of models of unbounded rationality as the gold standard (Klein, 2001). This issue is also applicable to Kahneman and Tversky’s general idea, as it contains the same implicit assumption that is worth noting. The conclusion that people make errors in reasoning can only be reached if there is a clear idea of what constitutes correct probabilistic reasoning first. Kahneman and Tversky unmistakably investigated heuristics in the light of the standards of rationality, and unsurprisingly found deviations. For them and others (e.g., Detmer, Fryback, & Gassner, 1978), who have conducted research on judgment heuristics, correct probabilistic reasoning is defined by Bayesian axioms, even though alternatives to Bayesian reasoning exist (e.g., Cohen, 1981; Gigerenzer & Goldstein, 1996, 1999). It is likely that behavior appears biased in one setting, but is well-adapted in another (Einhorn & Hogarth, 1981). Although the program of Kahneman and Tversky correctly argues that the way we make judgments does systematically deviate from the laws of logic or optimization, they hesitated to take a necessary further step – to rethink the norms set by the laws of logic and probability (Gigerenzer, 2006). However, the norms of logic and probability are not necessarily needed for making good decisions in a real world. So, if Dr. Plaster decided to get or stay employed, and therefore dismiss optimization or maximization as the gold standard, with what could he replace it?

¹³ The representativeness heuristic in its simplest form states that the probability of an event or outcome can be estimated by the degree to which it resembles the general population characteristics. The availability heuristic points out that different people arrive at different estimates, although having available objectively the same information due to a different background (memory). In the case of the anchoring and adjustment heuristic, former knowledge or experience with similar cases are used to serve as an anchor for the current case, after the anchor is set, adjustments to the current case are made.

2.2.12 *Do It Fast, Do It Frugal!*

Based on Simon's bounded rationality view, Gigerenzer and Goldstein (1996) developed an approach, which assumes that processing is *fast and frugal*. One of the key principles of this idea is that individuals adapt to their environments. Since there is only a certain amount of information that humans can process at any one time, it would appear rational for people to make use of the simplest form of information processing to reach acceptable decisions. So, in contrast to complex and compensatory models of decision making, it is proposed that people apply fast and frugal heuristics and still make sound decisions (Gigerenzer & Todd, 1999). Fast and frugal heuristics do not try to compute the maximum or minimum of some function, nor, for the most part, do they calculate probabilities (Gigerenzer & Kurzenhäuser, 2005). All of these fast and frugal models have in common that they invoke the smallest number of information (cues) and usually base their decision on one cue only, that is, they do not search through all the available information and they do not integrate all the relevant information. Many of these models are noncompensatory, so that the value on one cue cannot override the value on another cue, and are nonintegrative, as there is no real integration of information because the decision is often based on one critical cue. The models consist of principles for information search, for search stopping, and for decision making, so that, for example, a typical decision could be reached by searching through cues until the first discriminating cues are found. Different environments can have different specific heuristics that exploit their particular information structure to make an adaptive decision. Heuristics that employ one-reason decision making can exploit environments in which the importance of cues available is exponentially decreasing, that is, they are noncompensatory. In contrast, there are also heuristics that exploit environments that have a compensatory structure. For instance, the tallying heuristic does not employ one-reason decision making, however, its simplicity is that it uses only a few cues, does not order or weight them, and can search for cues in any order. In a study carried out on the decision when to send a patient into the coronary care unit instead of into the nursing bed, it was found that a simple decision tree by using only three questions performed equally well, compared to a complex regression model applying more than fifty (Forster et al., 2002). The repertoire of these different heuristics eligible for different environments has been named the *adaptive toolbox* by Gigerenzer & Todd (1999).

Such findings, which have their basis more on individual decision making, are fostered by research focusing on decision making in organizational groups, such as a team at a hospital would be. For instance, when faced with a problem to solve, Scholl (2003, 2004) found that decisions of organizational groups are not in line with the idea of being unbounded rational. Instead, he found the decisions of these groups often well-adapted to the environmental conditions, which he called *Adapted Rationality*, and what included the use of short cuts.

Although heuristics are meant to be well-adapted to the environment, they do not contain enough detail to match any one environment precisely, that is, their credit also lays in that they are not too specific, which ensures their accuracy. As heuristics consider only a few details and regard not every detail of paramount relevance, the risk of the occurrence of the phenomenon of *overfitting* is confined (e.g., German, Bienenstock, & Doursat, 1992). For example, while a model taking into account every single piece of information for describing a particular pattern might do better than a simple model, in terms of fitting, its predictive value for a comparable situation may be rather minimal, which, in fact, is rather useless in real-world situations. That is, by making use of only a little information, heuristics are not only well-adapted to our cognitive capabilities and environmental conditions but they are also characterized by robustness. The reasonableness of fast and frugal heuristics derives from their ecological rationality, not from following the classical laws of logic and rationality, such as internal consistency and coherence of choice. Fast and frugal heuristics can produce intransitive inferences in direct violation to the gold standard of rationality, but they still can be quite accurate (Gigerenzer, Czerlinski, & Martignon, 1999). How would Dr. Plaster decide on a pharmacodiagnostic test under the conditions of a fast and frugal mind? He might start with considering a first cue, the *severity of side effects of a treatment*, and, if severe, he would opt for the test, if not, he would consider the next one, *costs of the therapy*. If expensive, he would go for the test, if not, he would just discontinue search and decide not to order the test at all. This seems to be a more reasonable and practical decision-making model, given the usual 6-minute encounter between a patient and a doctor, than those models coming from the unbounded rationality side.

There is growing empirical evidence that physicians may rely on fast and frugal heuristics when they make their treatment decisions. A recent study carried out in the North of England on general practitioners and their hypothetical prescription for depression suggests that, with a fast and frugal model called Matching Heuristic (developed by Dhami & Ayton,

2001), doctors' decisions were almost as well explained as they were with a much more complex logistic regression model, but more economically, that is, by less cues (Smith & Gilhooly, 2006). Another study carried out in the same field (Smith et al., 2003) pointed toward the same findings. Dhami and Harries (2001) attained comparable results in their study conducted on general practitioners and their prescribing behavior for a lipid-lowering drug. Conditions, as outlined above, inherent in medical decision situations, such as task complexity as well as time pressure, have shown, in addition, to foster the use of simple decision strategies. Without question, heuristics appear to be powerful tools for medical judgments under the given conditions of complexity and uncertainty (Forster et al., 2002; Fischer et al., 2002). However, so far, studies conducted in this field, with the aim of comparing different models' fit and prediction to medical decisions, commonly focused on such decisions made by general practitioners or other clinicians, who are not confronted with potentially life-altering decisions. It is thus not clear at all if decision-making strategies of these health-care persons, being in charge of making such critical decisions, would also be comparably well described by a fast and frugal model than findings suggested from other professionals in the field of medicine. In addition, there is empirical evidence that general doctors are widely unable to understand and work with proposed concepts, such as the Bayesian formula recommended for deciding when to test. Would these findings also hold for clinicians in the field of cancer? Furthermore, if they do not require information in line with Bayes' theorem as taught, what type of information is it then that they use in order to make a decision on applying a pharmacodiagnostic test?

To answer these questions, a pilot study was conducted from Spring to Winter 2004 in both Germany and the USA, which dedicated the explorative character of these questions. In the course of the oncologists' pilot study, it became apparent that also the group of pathologists plays a role with respect to decisions on a pharmacodiagnostic test, and they play their own role. Therefore, a separate pilot study with pathologists was conducted during the same time period as the oncologists' study. Despite the exact title of being oncologists and pathologists, respectively, they are, and are seen by me as, just physicians, that is, everything outlined above about anything regarding decision making and probabilistic understanding holds true for them as it probably also does for most of us nonmedical human beings.

3 Study 1: Discovering Oncologists' Test Decision Making

“As soon as questions of will or decision or reason or choice of action arise, human science is at loss.”

Noam Chomsky

3.1 Pilot Study

3.1.1 Design

For the explorative nature of the pilot study, semistructured interviews with an open-answer format were conducted in order to learn about oncologists' decision making regarding the application of pharmacodiagnostic tests. The pilot study aimed at identifying relevant determinants of oncologists' decision making, finding out more about the environmental conditions and constraints of the decisions, as well as about decision-making strategies that oncologists would apply. That is, this should deliver sufficient insight in this area to generate a list of cues important for oncologists' decision making and to formulate adequate hypotheses.

Consequently, questions in the interview focused on how oncologists usually arrive at their treatment decisions without having any tests, whether or not they currently use any of the few tests already available, and if so, why they do so. Further questions addressed the hypothetical use of different other types of pharmacodiagnostic tests, since, at the time the interviews took place, only target tests and some crude prognostic tests were available to oncologists.¹⁴ Some additional questions were related to patients' influence on their diagnostic and treatment decision making in order to get a better insight of the patient's role concerning treatment decision making within the area of oncology. Table 3.1 summarizes each possible question.

3.1.2 Procedure

The participants were chosen at random from within the clinical areas of university clinics, hospitals, and private practices. Each oncologist was called upfront and asked for

¹⁴ In 2005, GenomicHealth launched the *OncotypDX*[®] test, which is a prognostic test.

Table 3.1: Catalogue of the Interview Questions for Oncologists

Topic	Questions
Demographic data of investigated person	Age Gender Qualification Position Years of experience Work location
1. Questions on experience	1.1 How do you usually make a decision on a suitable therapy? 1.2 Which criteria lead to your decision on the suitable drug? 1.3 Are there already any tests available to help you make a more appropriate decision about a drug? 1.3.1 If so: Could you explain, why do you use them? 1.4 When alternatives have been available, which criteria led to your choice? 1.5 Do you have any options in this respect (use/no use of tests)? 1.5.1 If so: Could you tell me what criteria lead you to make use of these options? 1.6 If not mentioned before: Does the accuracy or reliability of a test play a role in your decision? 1.6.1 If so: Do you have a preferred level for this? 1.6.2 Do you know, only roughly, the level of the current test used? 1.7 Could you give me an example of what “therapeutic consequence” means to you? 1.8 For some treatments there are tests available, which are optional. That is, it is your choice whether to use them or not. Could you explain, what criteria would you take into account in order to decide if you would use these tests or not? 1.9 Have you, in the past, used or considered using tests that are not covered by the medical insurances and/or not in the guidelines? 1.9.1 If so: Why did you decide to use them? 1.10 Can you remember any case where you have denied using a test? 1.10.1 If so: Why?
2. Questions concerning prognostic tests	2.1 Do you feel that this type of test is necessary to improve your treatment decisions? 2.1.1 If so: Why? If not: Why not? 2.2 Do you see a different necessity of tests of this kind depending on the cancer type? 2.3 Do you see a different necessity of tests of this kind depending on the treatment alternatives? 2.4 If such a test were available to you today, under which conditions would you opt for the usage of such a test? 2.5 Can you think of reasons against using such tests? 2.6 Could you also imagine features of a test that would stop you from using it? (sensitivity/specificity, communication of test results)
3. Questions related to patients’ influence on treatment decision	3.1 How do you consider the need of your patients participating in treatment decisions? 3.2 When you want to perform a test, do you ask/inform patients beforehand? 3.2.1 What information is normally given? 3.2.2 To what extent do you inform your patient about the test and its predicative value/error rate? 3.2.3 Are there any differences in the respective tests regarding how much information they yield? 3.3 Do patients ask for such types of tests?

Note: Not every participant was necessarily asked each question. It depended on the answers given for each question which questions were asked next.

their participation in this pilot study. Telephone numbers and basic information about each of the oncologists were obtained from the internet. Apart from one, each participant agreed to participate and was interviewed at their place of work. The interviews were tape recorded and lasted on average 1 hour (Germany, mean: 64 minutes; USA, mean: 59 minutes). Not every single question compiled for the pilot study was necessarily asked (see Table 3.1). It depended on the answers given for each question which question was asked next. When the interview was finished, each person received a book¹⁵ as compensation for their participation.

Table 3.2: Description of the Nineteen Participating Oncologists

3.2 Sample characteristics	Germany (12)	USA (7)
Age	40–64 years (Mean: 51 years)	40–63 years (Mean: 53 years)
Gender	Male: N = 10 Female: N = 2	Male: N = 5 Female: N = 2
Qualification	Medical oncologist: N = 10 Gynecological oncologist: N = 2	Medical oncologist: N = 5 Gynecological oncologist: N = 2
Position^a	Professor/ chief physician: N = 5 Associated professor/ Assistant medical doctor: N = 3	Professor/ chief physician: N = 3 Associated professor/ Assistant medical doctor: N = 3
Years of experience	7–34 years (Mean: 17 years)	9–34 years (Mean: 21 years)
Location/Institute	University/research clinic: N = 4 Hospital: N = 4 Private practice: N = 4	University/research clinic: N = 4 Hospital: N = 2 Private practice: N = 1

^a Oncologists practicing in a private practice do not hold a specific position in terms of hierarchy, as they usually have their own practice. That is, numbers do not add to the total of each sample due to this fact.

3.2.1 Participants

Twelve German oncologists, two women and ten men, practicing in Berlin, Potsdam, as well as in Leipzig, and seven US oncologists, two women and five men, practicing in Seattle (WA) were recruited as participants. Table 3.2 summarizes the characteristics for the

¹⁵ Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

German and the US sample. Due to time constraints for the US pilot study, not as many oncologists could be recruited for this study as for the German study.

For Germany, four oncologists were from university clinics, four participants were practicing in hospitals, and four in private practices. For the US sample, four oncologists were also located at university clinics, two oncologists in hospitals, and one in a private practice. The mean age of the German sample was 51 years (SD : 8.76), for the US sample this was 53 years (SD : 9.58). Oncologists' experience in treating cancer ranged for the German sample from 7 to 34 (mean: 17.50; SD : 8.20), for the US oncologists from 9 to 34 (mean: 21; SD : 9.92). The majority of these oncologists received training as medical oncologists and hematologists, that is, ten in Germany and five in the USA. The other two oncologists from each of the samples were gynecological oncologists. The US and the Germany pilot study sample were not significantly different from each other. With respect to the focused issue of this thesis, the samples from both USA and Germany are considered to be representative of the population of oncologists within each of the two health systems.

All participants were told that their answers would remain completely anonymous and confidential.

3.2.2 *Analytical Procedure*

All 19 interviews were transcribed and analyzed. Due to the explorative nature of the pilot study, the procedure chosen for the analysis was a simple one. Each interview protocol was assessed and checked for cues mentioned as being important for the decision on using such types of tests. Since the design considered for the main study set limitations on the number of cues to be taken into consideration, only those cues were extracted for the main study that were mentioned by at least 50% of each of the respective groups. The frequency of being mentioned was chosen as an indicator for importance, as it seems plausible to assume that things named by the majority of group members have a respective impact on their choice. It was also checked whether or not the pilot study results of both subsamples differed from each other in any respect.

3.2.3 *Results*

All oncologists interviewed had sufficient experience in treating tumors of different kinds. Concerning the employment of any pharmacodiagnostic test, all oncologists reported to

have sufficient experience with the current obtainable target tests. Some of participants (N = 6) also mentioned having applied the tests to guide their treatment decision, which was not recommended by the medical society at that point, but which was discussed as being potentially valuable. Hence, the pilot study participants were regarded as being suitably experienced to provide sufficient information with regard to the research object in question.

“Decisions on cancer treatment are pretty complex, which depend on balancing information along with experience and training. You know, each patient has his own history and deserves to be treated individually...”
A pilot study oncologist

What does drive the decision on using a pharmacodiagnostic test? Pilot study participants highlighted that the likelihood of an application of a pharmacodiagnostic test is strongly correlated with the treatment options at hand as well as the motivation to treat. If there was nothing to decide, a test would not be of use at all. What treatment options are finally available to the doctor and the patient is determined by the patient’s type and stage of tumor (see Chapter I) as well as the patient’s possible co-morbidities. Due to the resulting high degree of diversity of each of the cancer types, this usually leaves oncologists with no more than a few, if any, treatment alternatives, which is an issue important to stress. Since almost all of the currently available therapies are only modestly beneficial to the patients, oncologists want to make sure that they have utilized all of the few treatment opportunities that they have. The leading motivation is to make the patient survive, especially for the Stages I to III (see Chapter I), where a patient is assumed to have a certain chance of surviving the cancer. In this so-called “adjuvant treatment setting,” oncologists reported tending to overtreat the patients in order to extract every single survival percent of the drugs recommended for this cancer type and stage. For the palliative setting, Stage IV (see Chapter I), oncologists reported a different motivation regarding their style of administering, as they are confronted with patients who are assumed to die. The focus is shifted to the patients’ quality of life, that is, oncologists would be more likely not to fight the cancer by no longer giving extensive chemotherapy, but to relieve patients as much as possible from the burdens of cancer by only treating the effects of the disease itself.

Is a test needed at all? Indeed, there are some situations for which participants felt a test would be urgently needed. In the case of the adjuvant setting, a certain group of mamma carcinoma patients¹⁶ are usually treated with hormonal therapy and chemotherapy after surgery. However, several oncologists (N = 9 – for both) reported that for some 70% of this group, the so-called “good prognosis group,” it is assumed that they would survive¹⁷ the cancer by having only the hormonal therapy without having chemotherapy, which can cause severe cardiovascular side effects. This chemotherapy is helpful for another 10% of the patients of the hundred treated, who belong to the bad prognosis group. A quite comparable situation exists for the colon cancer population, where, for a certain group, a combination therapy is recommended. However, whereas one of the two chemotherapies benefits approximately 80% of the treated patients, the other therapy is beneficial for not more than 2% to 5% of the patients. The majority of the interviewed clinicians regarded having a prognostic test at hand, which could identify this subpopulation, are at good prognosis, and, therefore, could avoid overtreatment with unneeded therapies, as being tremendously helpful. Although both prognostic and responder tests (see Chapter I) have the potential to provide oncologists with information that would enable them to reduce overtreatment and consequently the dreaded side effects, most of them expressed their skepticism regarding having responder tests for the adjuvant setting. It seemed as if a more ethical concern was linked to this, as participants were worried that health insurances could exploit the treatment-specific information that such tests provide and reimburse only the treatments which had proven to be beneficial to the patient. Prognostic tests are not informative about whether or not a specific drug would be beneficial, and therefore may not have the potential of limiting treatment opportunities available to the oncologist.

For the palliative setting, where clinicians focus on the patients’ quality of life, the majority of participants, however, reported a desire to have responder tests which notify the doctor whether or not to administer a very specific treatment to the patient. It was highlighted that it is a difficult decision to make for both doctors and patients, as one would often gain only few months of life prolongation, if that, while accepting the suffering and pain from the side effects of administered therapies. Here, a responder test is assumed to be supportive in that it could tell whether or not the therapy itself is likely to deliver any life-prolonging

¹⁶ It is referred to the Estrogen Receptor positive (ER+) patients.

¹⁷ In terms of a 5-year survival rate.

benefits to a single patient, and, therefore, it might be worth enduring the side effects from the considered therapy.

If the situation would make a pharmacodiagnostic test appear helpful, the first thing reported to be considered is whether or not a test is recommended by the guidelines/ protocols of the oncologists' respective medical society¹⁸ or of their hospitals. Almost all of the participants emphasized that a recommended test would be more likely to be used in daily routine than one that is not. However, clinicians stated that guidelines are not obligatory. Most participants agreed that most of their past treatment and test decisions were rooted in guidelines and protocols to some extent, but asserted furthermore that the individuality of each patient still demands an enormous amount of unrooted and patient-specific decision making for which they apply their experience and training. Moreover, since guidelines are usually updated in intervals of 2 or 3 years only,¹⁹ those oncologists who were involved in research felt more updated about innovations within the field of oncology, and, therefore, ahead of the guidelines. For this reason, they regarded themselves to be more likely to apply tests even without having a recommendation by the guidelines as long as they consider the test helpful for their decision. In the case where the treatment would entail long-term severe side effects, it was furthermore reported by several participants that they were more likely to apply such a test with no recommendation by the guidelines.

Another cue mentioned as being influential on the decision of ordering a pharmacodiagnostic test was—unsurprisingly—the cost of the test itself. It is worth emphasizing that a recommendation in the guidelines usually accompanies the reimbursement of costs. Characteristics of a test, such as sensitivity and specificity, were rarely, if at all, mentioned by the participants themselves, let alone the predictive value. For those oncologists who did mention characteristics, more detailed follow-up questions were asked, which made it obvious that clinicians were barely able to work with these terms. Reasons given were a lack in understanding of this concept, and, therefore, an experienced inability to work properly with these terms. At this point of the interview, the majority of oncologists reported that they would hand over the responsibility for selecting a test by these criteria to their pathologists. As

¹⁸ There are several specific guidelines provided to the oncologists by disciplined specific organizations or the hospitals themselves. However, they usually do not differ remarkably from the general guidelines for Germany published by leading organizations, such as the German Cancer Society (DKG).

¹⁹ Guidelines are updated usually every 2 to 3 years; consequently, guideline recommendations may not concur with the most recent medical knowledge.

oncologists explained, it is their responsibility to decide whether or not a patient needs to have a test by answering a specific clinical question. If so, the oncologist would order the test from their pathological department, who is then responsible for selecting a suitable one. Interestingly, instead of sensitivity and specificity, oncologists frequently referred to a so-called therapeutic consequence that would have to accompany a test. This emerged as being a correlate of fractions of sensitivity and specificity, namely, the negative predictive value of a test. For the participants, a therapeutic consequence refers to the impact the test ultimately has on the treatment decision. That is, how many patients are spared unneeded chemotherapy (true negative rate of a test), and how many people are at risk of being undertreated by using the test (false negative rate). All interviewed clinicians mentioned the first consequence while fewer (N = 6 for Germany, N = 5 for the USA) seemed to be aware of the existence of the second consequence.

With respect to the oncologists' applied decision-making strategy, oncologists seem to go through a rather complex decision-making process, which includes several trade-offs. Following the reports of the interviewed participants, several cues related to the test itself, such as its recommendation, its cost, or its therapeutic consequence, which would be taken into account and weighed against each other as well as against cues related to the therapy.

Patients' influence on oncologists' treatment decisions. The final part of the interview was dedicated to the role of the patients' involvement within a treatment decision. Oncologists reported an increasing desire of patients to participate in the treatment decision. All oncologists were aware of the paradigm shift within society regarding patients' involvement, away from a paternal style of decision making toward a shared decision making (see Chapter V). This might have been presumably the reason why they stated that they would inform and involve their patients in any type of decision, including any kind of testing decision. Nevertheless, when asked in more detail, it was asserted that patients, in general, were not asked or especially informed when any type of test was performed on their tumor tissue, a procedure that is supposed to also underlie the majority of impending pharmacodiagnostic tests. For invasive test procedures, clinicians have to obtain permission due to legal regulations by their patients in order to execute the procedure. This means that patients are at least informed about it. In contrast to the German oncologists, US oncologists usually inform the patients about all test results after they obtain the results from the pathology department.

This was confirmed by the US patients' group (see pilot study of patients, Chapter V). German oncologists asserted that they deliver the test results that have an influence on the treatment decision. This was mostly not confirmed by the German patients' sample. In cases in which oncologists provided information, they delivered mainly general explanations about the testing procedure. The fact that tests deliver probabilistic information was not mentioned; this omission was ascribed to the patients' ignorance about testing. Regarding the final impact of patients on the treatment decision, oncologists referred to the strength of influence that patients could have on this decision, but in fact, patients quite rarely take advantage of this. That is, most cancer patients still believe in, and therefore follow, the recommendations given by their oncologists.

No remarkable differences between the German and the US sample were found that were related to the general decision situation that they reported to face in their daily practice, or to the decisive cues triggering and influencing their decision making. However, it should be mentioned that US oncologists reported a greater number of tests already used, and a higher level of patient involvement and impact on their treatment decisions, compared to their German colleagues. Table 3.3 summarizes the cues reported as being important for oncologists' decision making.

Table 3.3: *Cues Important for the Oncologists' Choice on Applying a Pharmacodiagnostic Test, Presented by Numbers of Mention*

<i>Cues</i>	<i>Germany (12)</i>	<i>USA (7)</i>
<i>Treatment options</i>	N = 12	N = 7
<i>Cancer stage</i>	N = 12	N = 7
<i>Cost of the treatment</i>	N = 9	N = 6
<i>Severity of the side effects of the treatment</i>	N = 10	N = 4
<i>Recommendation level of the tests</i>	N = 12	N = 6
<i>Cost of the test</i>	N = 9	N = 5
<i>Therapeutic consequence of the test</i>	N = 12	N = 6

3.3 Implications for the Main Study: Do Oncologists Have a Utility Function in Their Minds?

Following the reports of the interviewed participants of this pilot study, several cues related to the test itself would be taken into account and weighed against each other as well as against cues related to the therapy, which refers to a rather complex decision-making process, including several trade-offs.

I outlined in Chapter II that medical decisions are characterized by complexity, an abundance of probabilistic information, vague outcome feedback, and, furthermore, impacted by time pressure and reimbursement issues. In addition, in the case of deciding on whether or not to apply a test, complexity does even increase for clinicians, although none of the participants mentioned applying the Bayesian formula to decide on this issue. In contrast, it was found in this study that oncologists severely struggle with even basic concepts of test understanding, such as sensitivity and specificity. Given the complexity of the task as well as the other usual constraints mentioned earlier, I expect oncologists to be more likely to use a rather simple than a complex decision-making strategy when faced with a decision on a pharmacodiagnostic test. When looking at what participants reported when asked what had driven treatment decisions in the past, they frequently reported a recommendation of guidelines and protocols, respectively. Although studies (see Chapter II) point toward the fact that guidelines are quite often disregarded by clinicians, in the field of oncology where each decision is more a gamble with life given the severity of the disease, its diversity, and the mainly poor treatment options, for which benefits and risks are barely calculated according to the guidelines, may be regarded as helpful tools. For this study, that is, I expect oncologists' decision making to be equally well-explained by a fast and frugal model, as it would be by a complex model. In this respect, I assume that they would frequently apply the simple heuristic following the recommendation of the guidelines, as they also seem to do for other medical choices, such as their treatment decisions. This, however, was an assertion to prove. Therefore, the first hypothesis was to test whether or not oncologists would apply a simple heuristic when facing the choice of employing a pharmacodiagnostic test:

H1a: If the pharmacodiagnostic test is recommended by the guidelines and protocols, respectively, oncologists choose to use the test regardless of the other cue values.

However, with an increase in the severity of side effects, the impact of the recommendation cue on the choice is assumed to diminish, as especially long-term side effects in cancer treatments are recognized as making the administration of a treatment with an unclear benefit doubtful:

H1b: When the severity of the side effects of the treatment increases, more tests were chosen that have no recommendation in the guidelines.

It was furthermore highlighted that those oncologists who were involved in clinical internal research alleged to be guided by the guidelines/protocols to a lesser extent than those oncologists who were not. In addition, “research oncologists” who were involved in research emphasized significantly more use of complex decision-making strategies than oncologists who were not engaged in research. While I still assume to deal with human beings limited in cognitive capacity, and time and task is regarded as being complex, I nevertheless expected that the decision-making mechanisms that “research oncologists” would apply would not differ from those of the “nonresearch oncologists.” However, since I could see that this group did regard themselves as being, to some extent, ahead of the guidelines, I assumed to find the recommendation cue less impacting on their choices than on those of “nonresearch oncologists.” Of course, both of these assumptions, that is, a decreased impact of the recommendation cue and the call for a more complex decision-making process, also had to be proven. Therefore, the next two hypotheses relate to these issues:

H2a: The cue recommendation is expected to be found less impacting for those oncologists who are involved in clinical research, compared to oncologists who are not.

H2b: Research involved oncologists show no differences regarding the decision-making structure they apply, compared to their nonscientific counterparts.

Finally, another important implication of this pilot study was the appearance of the group of pathologists. The group of pathologists was not considered initially to be important with respect to the main question of this thesis. However, when coming to the topic of the

test's performance in terms of sensitivity and specificity, oncologists did refer to pathologists in charge for choosing tests on these criterion. As a consequence, pathologists were included in the general concept of the thesis. In Chapter IV I will outline the pilot study as well as the main study related to this group.

The rationale of the oncologists' pilot study was to learn and understand how oncologists would proceed when faced with a decision on whether to use a pharmacodiagnostic test for making a treatment decision. The findings of the pilot study enable me to determine important cues for this decision as well as to generate meaningful hypotheses. With respect to the general question, the rationale of the following main study was to obtain a comprehensive insight into oncologists' cue usage and decision-making strategies by applying compensatory, integrative as well as noncompensatory, fast and frugal models to the data. In this respect, it was not only of interest which of the two models might describe the data best (*comparison of the descriptive validity*) but also what model makes the best predictions for unknown cases, which would give details about the model's generalization (*comparison of the predictive validity*). Last, but not least, I was curious if I would find any intercultural differences that would show a more substantial and impressive effect than those found in the pilot study.

3.4 Dive Into the Bliss of Researching Human Judgment and Decision Making

Decision making in a range of applied domains including that of medicine (e.g., Wigton, 1996; Harries, Evans, Dennis, & Dean, 1996; Smith et al., 2003; Smith & Gilhooly, 2006) and occupational therapy (Harries & Gilhooly, 2003) has been analyzed extensively in recent years. With respect to the field of medicine there are some limitations, though. Since in the field of medicine each source of information has a probabilistic relationship to the true state of a patient and therefore the true relationship is not known, the majority of studies conducted in this field have been strategy-capturing ones only; that is to say, they focus on a description and prediction of clinicians' decisions (e.g., Dhimi & Harries, 2001; Harries et al., 1996), but are not able to tell how accurate a doctor's decision is with respect to the focused concern. As I faced comparable constraints, this study presented here also focuses on strategy-capturing and therefore I will not be able to tell how accurate clinicians' decisions are.

In order to capture a participant's applied decision-making strategy participants are typically asked to make decisions on a set of cases, real or hypothetical, which consist of a

combination of cues. Until recently, regression models were generally applied to infer which cues were being used and how cues were weighted in an assumed linear integration process (Cooksey, 1996). In this approach, participants' decisions are regressed on the cues presented, and the resulting equations represent the presumed judgment policies. These describe the amount and identity of information used (i.e., cues with significant weights), and how that information is weighted (e.g., standardized beta weights) and integrated (e.g., additive, compensatory rule). The majority of studies that have used this approach have found that a linear regression model fits individuals' judgment policies over a set of cases reasonably well, and that individuals' judgment policies contain no more than a handful of statistically significant cues (Dhimi & Harries, 2001). Although the regression approach has been a productive tool for investigating human judgment and decision making in applied domains, recently the use of regression models as descriptions of human decision making has been questioned (e.g., Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & The ABC Research Group, 1999; Dhimi & Harries, 2001; Dhimi & Ayton, 2001). Regression models are structural models. Due to this fact, these models provide a static description of judgment behavior only where the same information is used in the same manner when deciding on each case. Even though cue weights may be noncompensatory, and nonlinear terms may be included, it is generally assumed that judgments are a product of linear compensatory integration of multiple cues that are weighted optimally (Brehmer & Brehmer, 1988). However, as already stressed, the human mind is characterized by limited knowledge and by limited cognitive processing capacity (e.g., Kahneman, 1973; Miller, 1956; Olsson & Poom, 2005). In accordance with this research, it has been found that judgment strategies are chosen relative to the structure and demands of the judgment task (e.g., Hammond, 1996; Rieskamp & Hoffrage, 1999; see also Chapter II). Fox (1980) already found, by contrasting a probability-based normative theory of medical judgment (Bayesian probability revision) with a cognitive nonprobabilistic theory (set of heuristics and judgment algorithms), that both could predict the final diagnosis equally well. However, the latter gave a better account of the diagnostic process. Consequently, the psychological plausibility of static regression models for describing human beings decision making might be doubtful.

Gigerenzer and colleagues (1996, 1999) developed an approach derived from Simon's idea, which assumes that processing is *fast and frugal*. In contrast to regression models of decision making, it is proposed that people use rather very little information. All of these fast

and frugal models have in common that they invoke the smallest number of cues and usually base their decision on one cue only, that is, they do not necessarily search through all available information and they do not integrate all relevant information. The models consist of principles for information search, for search stopping, and for decision making, so that, for example, a typical decision could be reached by searching through cues until the first discriminating cue is found. Studies have compared the fit of fast and frugal models with compensatory integration models, such as regression models. Findings suggest that fast and frugal models describe judgment data at least as well as, and sometimes even outperform, compensatory models (e.g., Gigerenzer & Goldstein, 1996; Gigerenzer & Goldstein, 1999; Czerlinski, Gigerenzer, & Goldstein, 1999; Dhami & Ayton, 2001; Smith & Gilhooly, 2006). There have been some methodological concerns with such types of studies, however, since such comparisons are made on the basis of holding a multiple regression model that utilizes a continuous measure of judgment against a fast and frugal model that uses a binary measure of judgment. Nevertheless, in sum, fast and frugal models present a viable alternative to regression models. Inspired by the findings, a main study for oncologists was set up.

3.5 Main Study

3.5.1 Introduction

It was already hypothesized in the pilot study section that oncologists are assumed to apply to a proper amount a simple “rule of thumb” rather than a complex decision-making structure only such as described by them in the pilot study. The rule of thumb I expect them to apply is to simply look up the value of the cue *recommendation* for guiding their choice. Furthermore, intragroup differences are assumed regarding the degree of recommendation needed, but not regarding the applied decision-making mechanism. In addition to these hypotheses based on pilot study results, it is hypothesized that a fast and frugal model will comparably adequately describe and predict oncologists’ test decision-making policies as a compensatory integration model does. The hypotheses that they are likely to use simple strategies are based on the noticed environmental constraints, such as limited time, the availability of mainly probabilistic and incomprehensive information, as well as on the general limited computational capacity of human beings.

H3a: A fast and frugal model will provide an equally good fit to oncologists' test decision-making data as a compensatory integration model does.

H3b: A fast and frugal model will provide an equally good prediction to the unknown data of the oncologists' test decisions as a compensatory integration model does.

3.5.2 *Design*

The employment of discrete choice experiments, a judgment analysis method, was considered as being an eligible method for the main study. However, one of the problems in applying this method is that the evaluation of the problems of realistic complexity have a way of quickly generating a large number of multiattribute profiles if one persists on the use of a full factorial design. By considering just the seven cues for main study reasons, identified as the most important cues within the pilot study (see Table 3.1), each at three levels, a 3^7 design of 2,187 cases would result, given the desire for a full factorial orthogonal design, which is regarded as common in judgment analysis research (Cooksey, 1996). Apparently, the use of a full factorial design would yield an unmanageable number of cases. How then can these limitations be handled?

Ultimately, it was of interest to examine the effects of one cue on the decision independently of the effects of the other cues. This can be achieved at either the design stage using an orthogonal design or the analysis stage using, for example, partial correlations. A correlation-based analysis, such as the logistic regression, requires a large case to cue ratio (Tabachnick & Fidell, 1996), which had still meant that in the present study at least 106 cases had been needed to study the seven cues. Since pretests made obvious that clinicians would be willing to complete a maximum number of ten cases due to their lack of time, the decision was made to treat this issue at the design stage and to use a fractional factorial design that eliminated the possibility of testing most of the intercorrelations, but simultaneously retained the cues' orthogonality and ensured the verifiability of the main effects. For exploiting the restricted possibilities set by ten cases as much as possible, I opted for presenting the questionnaire randomly split in a version A and B in order to extend the number of cues to be manipulated. As a result of this design, I calculated that four cues out of the set of seven could be treated as predictors, while the constant cues were used to build the cover story for the

cases. Reasonably, it was decided to manipulate all of those cues that were related to the test itself, that is, *therapeutic consequence*,²⁰ *cost*, and *level of recommendation of the test*. Since four cues were allowed to be manipulated by the design, based on a discussion with some interview participants, the fourth cue chosen to manipulate was cue *side effects of the treatment*. The other two cues related to a description of the treatment, namely, *treatment options* and *cost of the treatment* as well as the cue describing the *stage of cancer*, were held constant. The design ensured that the intercorrelation between the test-related cues and the one cue related to the treatment were kept down to zero. The cue levels were specified in a consent process with three oncologists to ensure that cases would be as realistic as possible. This finally yielded a set of 18 cases, which were 9 cases per questionnaire version. Each of the cue values was equally distributed among the set of the 18 cases as well as among each version of the questionnaire. Table 3.4 shows the four cues and their related levels.

For the specification of the constant cues, a comparable counterpart of the treatment descriptions, which were reported within the interviews, was selected. Participants were presented with a situation that focused on the adjuvant setting (*cancer stage*), a therapy constellation consisted of therapy A plus therapy B, of which A already benefits 70% in terms of remaining disease free for at least 5 years, and therapy B an additional 5% out of 100 treated patients (*treatment options*). The *cost for therapy* was estimated at \$2,000 per patient for therapy A and \$10,000 per patient for therapy B. They were then asked to imagine that, for this specific situation, a test was available which would enable them to assign patients to either a good or a bad prognosis group. Depending on what the prognosis is, the patient needs to be treated with only therapy A to remain disease free (good prognosis) or requires therapy B in addition to therapy A in order to do so (bad prognosis). Then, the presentation of the nine cases of either version A or B of the questionnaire followed.

²⁰ For pilot study participants, a *therapeutic consequence* referred to how many patients are spared unneeded chemotherapy (true negative rate of a test), and how many people are at risk of being undertreated by doing the test (false negative rate). Therefore, it was decided to manipulate these two facts within the presented cases. In case one of the main study, oncologists should also wish to know values for the true positive and false positive rate, a clickable overview (see Appendix A) was included on the webpage providing all values (positive, negative, false, true) in a natural frequency format as well as in probabilities. For two reasons, I deemed this procedure to reflect the reality best: First, if oncologists desire sufficient information on the quality of the test in their daily practice, they would have to make some extra effort to be provided with such information. Second, several pilot study oncologists seemed not even to know about the existence of the negative rates. By having these put into the cases, I had drawn participants' attention to this matter, which they would never have thought about under real conditions.

Table 3.4: *Manipulated Cues, Their Levels and Distribution for Versions A and B of the Questionnaire*

Cues	Level	Distribution	
		Part A	Part B
Severity of the side effects of the treatment	(0) Permanent serious side effects	(0) 4	(0) 5
	(1) Serious side effects during treatment	(1) 5	(1) 4
Therapeutic consequence of the test²¹	(0) 350 patients would be spared overtreatment (out of 700/1000), 2 patients would be undertreated (out of 50/1000).	(0) 3	(0) 3
	(1) 350 patients would be spared overtreatment (out of 700/1000), 4 patients would be undertreated (out of 50/1000)	(1) 3	(1) 3
	(2) 350 patients would be spared overtreatment (out of 700/1000), 6 patients would be undertreated (out of 50/1000).	(2) 3	(2) 3
Cost of the test	(0) €/\$ 300	(0) 3	(0) 3
	(1) €/\$ 800	(1) 3	(1) 3
	(2) €/\$ 1,300	(2) 3	(2) 3
Level of evidence of test	(0) Recommended by your respective medical society and guidelines (costs covered by health insurance)	(0) 3	(0) 3
	(1) Recommended by some of your experts, but not by your respective medical society or guidelines (costs covered by health insurance)	(1) 3	(1) 3
	(2) Neither recommended by your experts nor by your respective medical society or guidelines (costs covered by health insurance)	(2) 3	(2) 3

For each case, oncologists were asked to give a yes/no response as to whether or not they would apply the pharmacodiagnostic test under the described conditions. Since the involvement of patients in the medical decision (shared decision making) has become increasingly expected in recent years, the “Yes, I would use the test” was linked to “or recommend it to the patient respectively” to give participants of the main study the feeling that

²¹ On the respective webpage, oncologists were provided with a clickable overview that gave detailed and comprehensive information not only regarding the true positive rate and the false positive rate of the test, as already displayed in the cases themselves, but also information on their true negative rate and the false negative rate. Each piece of information was delivered in a natural frequency format as well as in probabilities.

a yes vote did not necessarily mean that they were patronizing their patients by ignoring their preferences. For the cover story, it can be stressed that no particular patient was described to the participants, but they were asked to imagine that they already consider the administration for the particular treatment combination A and B. This was to reduce the well-recognized treatment variance (Wigton, 1996) between clinicians. Since a detailed description of a patient had fostered different ideas of what treatment would be an appropriate one for that patient, which in fact was of no interest for this study but the test decision itself, I hoped to keep such types of biases out of the design by being already prescriptive for the treatment under consideration. After presenting the nine choices, five questions concerning the questionnaire's representativeness were asked to find out how straightforward the described situation and cases appeared, whether or not answering the questionnaire was felt to be easy, if participants felt they had received sufficient information to make a reasonable choice, and whether or not they missed cues by doing so. Finally, I asked the participants how long it took them to work through the nine choices. The respective questions are shown in Table 3.5.

Table 3.5: Questions Concerning the Quality of the Questionnaire

Question	Values
F1: I found answering the questionnaire:	(1) Easy from start to end (2) Easy at the beginning, more difficult toward the end (3) Difficult from start to end
F2: The situation/setting described in the questionnaire was:	(1) Straightforward (2) Sometimes abstract (3) Rather or totally abstract
F3: The decision criteria generally offered in the scenarios (e.g., experts' recommendations, cost of the test kit) would also be criteria I would use in medical decision making:	(1) I agree (2) I partly agree (3) I do not agree
F4: Did you feel one or more decision criteria were missing within the scenarios that you would normally use in medical decision making?	Open answer field
F5: How long did it take you to fill out the questionnaire?	Open answer field

Furthermore, questions concerning participants' age, position, work location, involvement in research, as well as the frequency of attending scientific meetings or education training were included, as they had shown, to some extent, to be influential on choice within the pilot study. These questions with their values are outlined in Table 3.6.

The main study task was piloted on four German and three US participants from the pilot study, and only little adjustment on more precise background information was made.

Table 3.6: *Questions and Their Values for Capturing Intragroup Differences and Demographics*

<i>Questions</i>	<i>Values</i>
D1: How old are you?	(1) < 35 (2) 35 - 50 (3) > 50
D2: What is your exact job title? (e.g., Med Onc, Int Med, Med Onc/Hem)	Open field answer
D3: How many years have you practiced as a professional?	Open field answer
D4: Where do you work?	(1) Academic setting/university clinic (2) Hospital (3) Private medical practice
D5: When you work in a clinic or hospital, which of the nominations below do describe your position best?	(1) Chief physician/head of department; professor (2) Assistant medical director; assistant professor (3) Assistant doctor; staff physician
D6: Do you conduct your own research projects in your academic or private practice? (except clinical trials)	(1) Yes (please continue with question D7) (2) No (please continue with question D8)
D7: Are you involved in any of these research projects?	(1) Yes, as principal investigator (2) Yes, as a co-investigator (3) No
D8: How many times a year do you attend scientific meetings of your medical society/community?	(1) Less than three times a year (2) Three to five times a year (3) More than five times a year
D9: How many times a year do you attend professional continuing education?	(1) Less than once a year (2) At least once a year (3) At least twice a year (4) More than twice a year

3.5.3 Procedure

A detailed description of the medical situation in question and the related nine cases were presented on a webpage to the participants. Upfront, a short foreword was included, which introduced the study as well as the investigator, and highlighted and guaranteed respondents anonymity. At the end of the webpage, questions considering any intragroup differences and demographic data as well as questions regarding the quality of the described medical situation and the related nine cases were included (Appendix A). Either Version A or B was presented randomly to each of the participants, whereby only the cases presented, but not any other part of the questionnaire, such as the description of the medical situation or demographic questions, were varied. As the chosen medical situation focused on a decision to

make on chemotherapies for a solid tumor patient, only such oncologists who were trained in this area were defined as being eligible for participating in this study. In Germany, several organizations, which are regarded as having access to representatives of this oncologist community, were asked for their support by distributing a mail to their members,²² which provided a short explanation of the investigator and her research, the request for participation, as well as a link to the respective webpage (see Appendix A). The following societies and associations finally provided their members with this email: German Network of Evidence-Based Medicine (ebm), the German Cancer Society (DKG) with their related Consortium of Internal Oncologists (AIO), the Consortium of Gynecological Oncologists (AGO), as well as the Consortium of Urological Oncologists (AUO); Working Committee of German Tumor Centers e.V. (ADT), Northeast German Society of Gynecological Oncology e.V. (NOGGO), and the German Society of Hematology and Oncology (DGHO).²³ Since not all societies or associations, respectively, revealed the number of their members, the response rate for the German sample could only be estimated at approximately 10%, which is rather low. However, when no type of professional engagement between the researcher and the participants exists, as was the case for this study, such rather low response rates seem to be quite usual for this group of medical experts. The sampling frame for the US oncologist sample was obtained from the American Society of Clinical Oncology (ASCO) member directory 2005. Over 2,500 medical and gynecological oncologists were approached directly via email by the investigator and asked for their participation in the study. The content of this email corresponded to the one distributed by the German societies to their members (see Appendix A). The final response rate was approximately 5%, which is only half as many as for the German sample. However, this is considered an effect of approaching US oncologists as a private person and not as one of their respected societies, as was the case in Germany.

²² They were explicitly asked not to reveal their member lists to the investigator to ensure the highest level of privacy to the participants.

²³ Netzwerk evidenzbasierter Medizin (ebm), Deutsche Krebsgesellschaft (DKG), Arbeitsgemeinschaft Internistischer Onkologen (AIO), Arbeitsgemeinschaft Gynäkologischer Onkologen (AGO), Arbeitsgemeinschaft Urologischer Onkologen (AUO), Arbeitszentren deutscher Tumorzentren e.V. (ADT), Nordostdeutsche Gesellschaft für Gynäkologische Onkologie e.V. (NOGGO), Deutsche Gesellschaft für Hämatologie und Onkologie (DGHO).

3.5.4 *Participants*

One hundred and eleven German oncologists and 109 US oncologists fully completed the online questionnaire. As the initial invitation for participating in the study was distributed by email to oncologists across both countries, participants in each region of both countries should have been captured.

Table 3.7: Demography and Intragroup-Related Values of the Main Study Oncologists

Questions	Values	Germany (N = 111)	USA (N = 109)
D1: How old are you?	(1) < 35 (2) 35–50 (3) > 50	N = 7 N = 71 N = 33	N = 11 N = 61 N = 37
D2: What is your exact job title? (e.g., Med Onc, Int Med, Med Onc/Hem)	Open field answer	Medical oncologist/ Internal oncologists: N = 83 Gynecological/ urological/ hematological oncologist: N = 28	Medical oncologist/ Internal oncologists: N = 86 Gynecological/ urological/ hematological oncologist: N = 23
D3: How many years have you practiced as a professional?	Open field answer	Range: 4 – 33 (mean: 17,95) 1 missed value	Range: 2 – 40 (mean: 15,84)
D4: Where do you work?	(1) Academic setting/University clinic (2) Hospital (3) Private medical practice	N = 41 N = 50 N = 19 1 missed value	N = 58 N = 19 N = 31 1 missed value
D5: When you work in a clinic or hospital, which of the nominations below do describe your position best? ^a	(1) Chief physician/Head of department; Professor (2) Assistant medical director; Assistant professor (3) Assistant doctor; Staff physician	N = 25 N = 49 N = 17 1 missed value	N = 38 N = 33 N = 18
D6: Do you conduct your own research projects in your academic or private practice? (except clinical trials)	(1) Yes (Please continue with question D7) (2) No (Please continue with question D8)	N = 43 N = 67 1 missed value	N = 80 N = 39
D7: Are you involved in any of these research projects?	(1) Yes, as principal investigator (2) Yes, as a co-investigator (3) No	N = 25 N = 18 N = 67 1 missed value	N = 50 N = 26 N = 33
D8: How many times a year do you attend scientific meetings of your medical society/community?	(1) Less than three times a year (2) Three to five times a year (3) More than five times a year	N = 7 N = 54 N = 50	N = 41 N = 41 N = 27
D9: How many times a year do you attend professional continuing education?	(1) Less than once a year (2) At least once a year (3) At least twice a year (4) More than twice a year	N = 2 N = 9 N = 14 N = 86	N = 2 N = 21 N = 18 N = 68

^a Oncologists practicing in private practices do only rarely hold a specific position in terms of hierarchy, as they usually have their own practice, especially in Germany. Therefore, numbers do not add to the total of each sample.

Both the German and the US sample of oncologists who participated are considered to be representative of the community of oncologists who are in charge of making decisions on

chemotherapy treatments for cancer patients with a solid tumor, and who are, therefore, likely to face a decision on a pharmacodiagnostic test in their practice.

To start with the German sample ($N = 111$),²⁴ oncologists' experience in treating tumors ranged from 4 to 33 years, with a mean of 17.95 years ($SD: 7.89$). Nineteen practiced in private practices, 50 in hospitals, and 41 in university clinics. Twenty-five of the German oncologists were chief physicians/professors, 49 assistant medical doctors/associated professors, and 17 assistant doctors/staff physicians. Forty-three reported being involved in research, while 25 were principals, and 18 coinvestigators.

For the US oncologist sample ($N = 109$), their experience in treating tumors ranged from 2 to 40 years, with a mean of 15.84 years ($SD: 8.89$). Thirty-one worked in private practices, 19 in hospitals, and 58 in university and research clinics, one value was missed. The majority of the US participants, namely 38, were chief physicians/professors, followed by 33 assistant medical doctors/associated professors, and 18 assistant doctors/staff physicians. Of the 109 US oncologists, 80 were involved in research, while 50 were research principals, and 26 coinvestigators.

Significant differences between the German and US sample were found regarding their work location, the research involvement, as well as the frequencies of attending meetings and educational training. These differences are probably based on the different recruitment procedure of the US study. Here, only oncologists were approached whose addresses were collected from an annual scientific meeting member directory of ASCO. That is, such physicians were contacted who were already more likely to attend a meeting, which, in addition, is also regarded to be a type of educational training due to its exclusive presentations and workshops there. Table 3.7 summarizes all data.

3.5.5 Analysis Considerations

I aimed to compare a noncompensatory fast and frugal model with a compensatory integration model in order to find out which of these models fits and generalizes oncologists' test decisions best. Logistic regression usually is chosen for analyzing data, such as those obtained by this study. However, as I was interested in studying decision-making on the level of individuals, the case-to-cue ratio presented to each participant (i.e. 3:1) in this study was too

²⁴ Within the German sample, for one participant the value from question D3 to D7 was missed, therefore, numbers do not always add up to 111 participants.

low as to being able to apply this model to the data without risking getting distorted results (Tabachnick & Fidell, 1996). Therefore, two other compensatory integration models with different degrees of complexity were chosen and individually applied to each oncologist's data instead. I decided to use a differentially weighted integration model called *Franklin's rule* and a unit weighted integration model called *Dawes' rule*, both of which were also used by Gigerenzer and his colleagues (1999). The advantage of these two models is that they provide characterizations of judgment behavior similar to that provided by a regression model. Although both models do not weight the cues optimally in the way the least-squares method does in a regression model, several studies have demonstrated that both models are first-rate approximations to regression models in terms of descriptive and predictive validity (Cattin, 1978; Dawes & Corrigan, 1974; Dorans & Drasgow, 1978; Einhorn & Hogarth, 1975; Czerlinski et al., 1999; Gigerenzer & Goldstein, 1996; Schmidt, 1971; Wainer, 1976; Dhami & Ayton, 2001). In this sense, I do not consider the present study to be especially limited due to the inability to compare a noncompensatory model with a regression model.

For the oncologists' main study, as a competitor on the noncompensatory side, the *Matching Heuristic* (MH) model by Dhami and Ayton (2001) was chosen, as the binary categorization decision task in their research was very similar in form to the present one.

It is worth highlighting, however, that in contrast e.g., to Gigerenzer's research (1996, 1999) where inferences were studied about several real world aspects for which the actual external validity can be calculated (e.g., which of two cities is larger), I was not able to determine an outside-criterion with respect to whether the choice made is correct or not. In case of medicine it is often quite difficult to model both the judgment and the ecology (Speroff, Connors, & Dawson, 1989; Wigton, 1996) as cue weights in the ecology are not known for the majority of diagnostic and treatment judgments (Wigton et al., 1986; Wigton, 1996). Weights are difficult to obtain and are often situation-dependent (see Chapter II and previous paragraphs within this chapter). Therefore, in analogy to the commonly reported cue validity for models such as Franklin's rule, I determined each cue's importance. The importance was defined by the frequency of how often a respective cue value was followed by a choice.

For ease of analysis of models' fit and prediction, polytomous cues were dichotomized. For each cue, all italicized values were coded as "0" and nonitalicized values were coded as "1" (see Table 3.4). The dichotomization was based on weights derived from an overall

multiple regression carried out on the overall data per country group. Although there was already information about what cue levels would rather lead to a “yes” or “no” choice due to the pilot study, I asked for a further and more statistical rationale. For this reason, cue values of each cue were dummy coded, for example, in the case of a tripartite cue either the “best level” and the “medium level” were coded “1” while the “lowest” was coded “0” for dummy variable condition 1, or for dummy variable condition 2 the “best level” was coded “1” only while the “medium” and the “lowest level” were coded “0.” In cases where the cue was only of binary nature, which was the case for *side effects of the treatment* only, no dummy code was applied, as cues had been already coded priorly in a “0”/ “1” manner. The dummy variables were used as an independent variable to predict the proportion of all “yes” choices for each case. For this reason, the proportion of the “yes” choices for each case over all choices per country was calculated, and then it was tested whether the dummy variables of condition 1 or those of condition 2 predicted the proportion better. This was done for all of the 18 cases. In this way, it was possible to define whether the medium level of a tripartite cue was more likely to cause a “no test” choice or rather likely to trigger a “yes test” choice. Based on the result, which did not have to be statistically significant but just had to have a higher predictive value, the medium level was sorted either to the group of the high level of a cue or to that of the low level of a cue. The finally derived dichotomized cues were used for all of the following models applied on the oncologists’ data set, Franklin’s rule, Dawes’ rule, and Matching Heuristic.

As the choice was already of binary nature, there was no need to simplify this. The test decision for each oncologist was modeled on a set of 9 cases.

In the following, I will first describe the procedure for modeling the oncologists’ test decisions with each of the three models (Franklin’s rule, Dawes’ rule, and Matching Heuristic). For all these models, the default was the “no test” decision. After outlining the models’ procedures, the results of the fit of each model as well as the performance of each of its generalizability to unknown data will be presented, followed by additional results. This is done separately for the German and the US sample. The model with the best prediction in terms of percentage of correctly predicted decisions of each single participant, in the following referred to as *best average prediction*, was chosen as the model of the oncologists’ test application policy. For comparing fast and frugal models with linear models, the index of fit

*percentage correct predictions*²⁵ has been used, as common metric in previous studies (e.g., Czerlinski et al., 1999; Martignon, 2001).

Franklin's Rule. This model is certainly very close to the idea of unbounded rationality, and within this model competition regarded as the most complex one. In this model, each cue was weighted according to its influence on the decision. Then, for each case, this model multiplied the cue values by their weights and then summed them. In such a case as ours, where binary cues are coded 0 and 1, the sum of the cue weights is the sum alone for all those cues taking a value of 1 in the respective case, which is exactly comparable with how the regression model equation would proceed. If the sum was equal to, or greater than, the threshold value then a “yes test” decision was predicted, if not, a “no test” decision was predicted.

The cue weights were calculated by choosing the value on each cue that had the greatest proportion of “yes test” decisions in the set of 9 cases. For example, if the proportion of “test recommended” led to a higher proportion of “yes test” decisions than “test not recommended,” let us say 90% to 10%, then the former one would define the weight (.90) for the recommendation cue. For the model to make a decision on whether or not to predict a “yes” choice, it needed a threshold. The threshold value was calculated first by taking the sum of the calculated cue weights for each of the nine cases, totaling the sum of these nine cases, and then dividing this total by the number of cases. After the cue weights and the threshold were defined, for each of the nine cases, the model was asked to predict what the outcome per case would be. If the predicted outcome was the same as the actual oncologist's choice, a hit was recorded. After all nine cases were examined, the hits were summed and set in relation to the nine cases which led to the final percentage of fit for that participant. Furthermore, the model's ability to predict unknown cases (generalization) was investigated by using the *k*-1 cross-validation technique. If the prediction was correct, a “hit” was recorded. To provide a concrete example, I used the oncologist-1's test decision-making policy for the first eight choices, as described by Franklin's rule, to predict this oncologist's decision in the ninth case. In this case, the treatment had *permanent serious* side effects, *four people* were undertreated due to a false test result, test's cost were €1,300, and the test was *not recommended* by respective guidelines. The weights attached to the cues were as follows: *side*

²⁵ Please be aware of the fact that the participants of the oncologist study were faced with a binary “yes”/“no” choice, that is, only one object of choice is present and not two or more. Therefore, each cue's importance weight is not based on discriminative values of a cue, as it would be the case when at least two objects of choice were presented (see Chapter IV).

$effects(1)(.66)+under-treat(0) (.66)+test's\ cost(0)(.66)+recomm(0)(1.00)$. The sum is .66, which is less than the 1.13 threshold value calculated for that oncologist. Thus, in this case, Franklin's rule would correctly predict that oncologist-1 made a "no test" decision.

Dawes' Rule. Dawes' rule is still a compensatory model, but uses unit weights, and is therefore less complex than Franklin's rule. It simply works by adding up the number of positive cue values and subtracting the number of negative cue values. Thus, it does not involve much computation and due to this it is fast, but still looks up all cues, but is therefore not frugal (Gigerenzer & Goldstein, 1999).

For the analysis, the model only evaluates whether or not a cue had a positive (critical) value for the choice or not. For each case, this model counted how many cues the critical cue value had, and if the sum of these critical cue values was greater or equal to the threshold value, the model would predict a "yes test" decision. The critical value of the cue was defined as the value with the greatest proportion that had led to a "yes test" decision in the set of nine cases. For example, if the proportion of "test recommended" led to a higher proportion of "yes test" decisions than it did for "test not recommended," then the critical cue value for the recommendation cue would have been "test recommended." This critical cue value was then given a weight of "1," while the other cue value was given a weight of "0." The threshold value was again calculated by first taking the sum of the cue weights for each of the nine cases, totaling these nine sums, and then dividing the total by the number of cases (i.e., nine). Also here, the threshold was for the model to decide whether or not to predict a "yes" or "no" choice.

Dawes' rule's fit was defined by its ability to predict the actual outcome for each of the nine cases. If the predicted outcome was the same as the actual oncologist's choice, again a "hit" was recorded. After all nine cases were predicted by the model, the hits were summed and set in relation to the nine cases, which led to the final percentage of fit for that participant. As was already outlined in Franklin's rule, furthermore, the model's ability to predict unknown cases was of interest and investigated in the same way as it was for Franklin's rule. To also provide an example of the Dawes' rule's ability to predict oncologist-1's unknown ninth choice based on a model derived from the first eight cases: $side\ effects(1)+under-treat(1)+test's\ cost(1)+recomm(0)$ yield a sum of 3, which is more than the 2.00 threshold

value calculated for that oncologist. Dawes' rule would incorrectly predict that oncologist-1 made a "yes test" decision.

Matching Heuristic (MH). The MH model proposed that decision makers search through a certain number of cues until the first critical cue is found, and then they stop and decide, for example, to test or not to test. That is, the MH model assumes a noncompensatory decision-making process. The maximum number of cues an individual searches through is K and the cues are searched through in order of their importance. A person might search through one, two, or more cues depending on when a critical value is found, and subject to what K is for that individual. If no critical value was found after K cues had been examined, then the default response is made, which in the present oncologist study would be "not to test," but adhere to today's procedure of administering the treatment without upfront testing. An example of the model could be such as the one shown in Figure 3.1. Here, an oncologist would employ a $K=2$ model to make their decision ($K=2$). In this example, the two cues used were "cost of the treatment" and "severity of side effects of the treatment." When deciding on a pharmacodiagnostic test, the oncologist would first look at the "cost of the treatment" (given that this is the most valid cue for this oncologist) and if it has the critical value, which is in this example presented as "1," then they would apply a test. However, if the value on this cue is "0," that is, does not have the critical value, they would consider the second most valid cue, for example, "severity of side effects of the treatment." If this cue has the critical value "1," then the test is applied, otherwise it is not. This would be the end of the procedure.

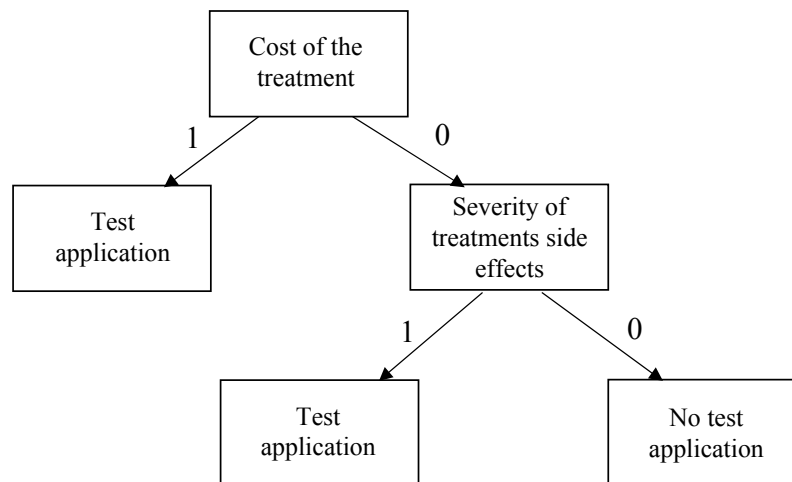


Figure 3.1: Example of a flowchart of Matching Heuristic searching through a maximum of two cues ($K = 2$ model).

For determining the fit of the MH model to the oncologists' data, the following three steps were taken. First, the critical value of each cue was found for each participant. The critical value of each cue is the value of a cue which is most often followed by a "test use" decision. If there were more "1s" than "0s" for a given cue on the "test use" decision, the critical value of that particular cue would be "1." If there was a tie between "0s" and "1s," then a critical value would be chosen randomly. The second step was to find the utilization importance for each cue. This is defined as the proportion of cases with the critical value that had a "test use" decision. So, for example, if for a particular oncologist the "severity of treatments side effects" cue had the critical value of "1," and in 70% of cases with a value "1" on that cue the oncologist made the "test use" decision, the cue would have a importance of 0.70 for the oncologist concerned. The cue validities were then used to rank order the four cues, where the first ranked cue had the highest importance. This rank order is the order in which cues are added to the model in order to determine K, the maximum number of cues to be searched. When cues had tied validities, the rank order was determined randomized. The third and final step involved finding K, the maximum number of cues used by each oncologist. This was determined by starting with a one-cue model, using the most valid cue, and determining how well the individual oncologist's decision is predicted by the use of that single cue. Then the whole procedure was repeated by using a two-cue model, and so forth, and was stopped if including another cue led to no improvement or even reduced the fit. Unlike regression models, adding more cues to the model does not necessarily result in an improved fit. As there is no significance testing of the degree of improvement, for that study any absolute improvement led to the relevant cue being added. With nine cases, the smallest improvement was 11%, that is, one additional case is predicted. In this way, K was determined for each individual oncologist. The fit of the Matching Heuristic model for each oncologist was the proportion of correctly predicted cases using a K-cued model for that participant.

Furthermore, the model's ability to predict unknown cases (generalization) was investigated in the same manner as it was for the other two procedures. If the prediction was correct, a "hit" was recorded. To provide a concrete example, I used the oncologist-1's test decision-making policy for the first eight choices, as described by Matching Heuristic, to predict this oncologist's decision in the ninth case. In this case, the model ranked the cues as following: *recomm(1)*, *side-effects(1)*, *under-treat(1)*, and *test's cost (1)*. As for the ninth case,

the cue *recommendation* did not have the critical value, the model predicted correctly a “no” choice for this case by only using a $K = 1$ model.

3.5.6 Results

In the following, I will first give a short overview on the main study participants’ overall choice behavior before I start to dwell on answering all hypotheses posed under the previous main study section as well as under the pilot study section.

Should I say “yes” or “no”? Oncologists overall choice and agreement between them. Overall, in 65.9% of the cases, oncologists decided to order a pharmacodiagnostic test, while within the German sample 60% did so and within the US sample 72% of the participants. Over the set of nine choices, German oncologists opted for a test 5.4 times, on average, while their American colleagues did so 6.5 times. This difference between the two countries was significant ($t(218) = -4.05, p = < .001, r = .26$) and of small size (Field, 2005). Taking a closer look at each sample, for the German one, the majority who opted for a test were oncologists coming from the university (65.3%—mean over the set of nine choices: 5.88), followed by oncologists working in private practices (57.3%—mean over the set of nine choices: 5.12), and by hospital oncologists (56.9%—mean over the set of nine choices: 5.16).

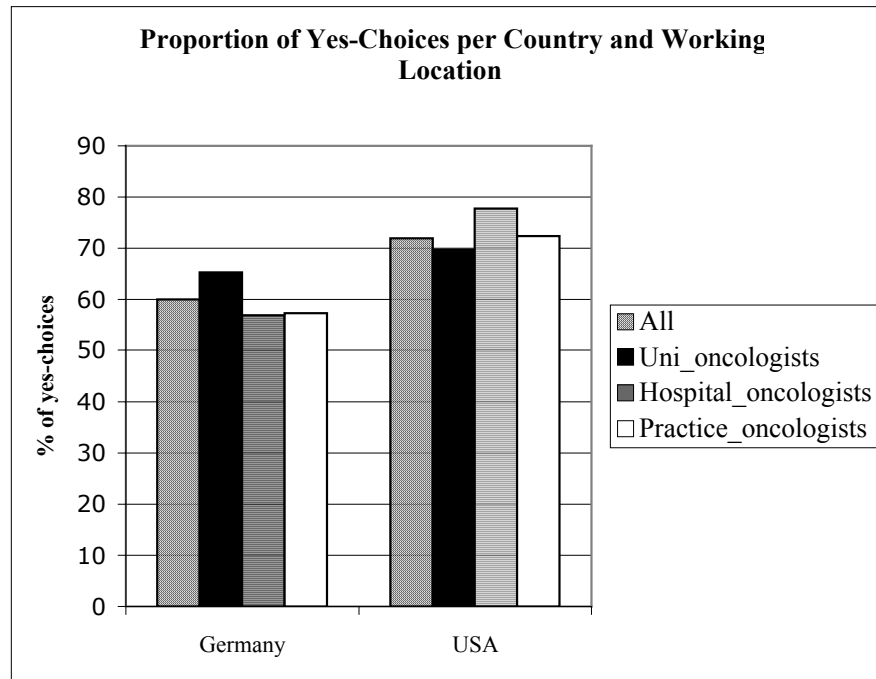


Figure 3.2: Proportion of yes-choices per country for all oncologists (*All_oncologists*) as well as separated between oncologists coming from university clinics (*Uni_oncologists*), hospitals (*Hospital_oncologists*), and private practices (*Practice_oncologists*).

The picture was completely reversed for the US sample, here hospital oncologists opted mostly for doing the test (77.8%—mean over the set of nine choices: 7.00), with oncologists from private practices again in a middle field position (72.4%—mean over the set of nine choices: 6.52), and finally, oncologists who worked at university clinics (69.7%—mean over the set of nine choices: 6.28). No differences regarding the proportion of opting for a test were found within each of the samples ($F(2,107) = 1.81, p = .169, \omega = .13$; $F(2,105) = .993, p = .374, \omega = .01$). Figure 3.2 exhibits data per country and per location of work. When investigating differences of the yes-choice behavior between the three groups of each country, no difference was found for the university groups of the two countries ($t(97) = -1.01, p = .387, r = .10$), while a significant if small difference between the two hospital groups ($t(67) = -3.25, p = .002, r = .13$) and a medium sized difference for the private practice groups ($t(48) = -2.58, p = .013, r = .35$) was found.

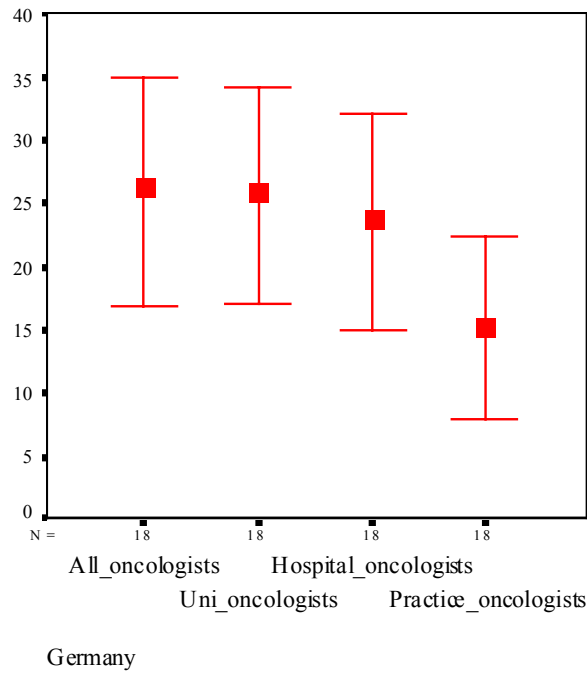


Figure 3.3: Disagreement over all 18 cases (9 per version) of the German oncologist sample (mean: 95% CI) for all oncologists (All oncologists) as well as separated between oncologists coming from university clinics (Uni oncologists), hospitals (Hospital oncologists), and private practices (Practice oncologists)

In order to measure the extent of consistency across participants, the percentage of oncologists' disagreement with the model choice on each case was calculated. The modal choice is defined by the choice outcome of each case most of the participants opted for. Within each sample, there was disagreement between oncologists regarding the decision to be made on each of the nine cases. The disagreement ranged within the German sample from 1.8% to 48.20% (mean: 26.00; *SD*: 18.12), and for the US sample fairly identically from 1.8% to 49.10% (mean: 21.30; *SD*: 17.15). This difference in disagreement between the two countries was not significant ($t(34) = .799, p = .433$). German oncologists' disagreement, who came from the university, ranged from 0% to 47.37% (mean: 25.70; *SD*: 17.42), while for their US counterparts this ranged from 0% to 47.62% (mean: 20.94; *SD*: 16.90). Regarding the oncologists coming from hospitals, disagreement between those coming from Germany ranged from 4% to 48% (mean: 23.56; *SD*: 17.40) and within the US sample from 0% to 43.45% (mean: 17.69; *SD*: 18.60). For oncologists from private practices, in Germany they disagreed ranging from 0% to 41.67% (mean: 15.15; *SD*: 14.37), while this was from 0% to 50% (mean:

17.87; *SD*: 14.95) in the US sample. All values for Germany can be seen in Figure 3.3, and for the USA in Figure 3.4, respectively.

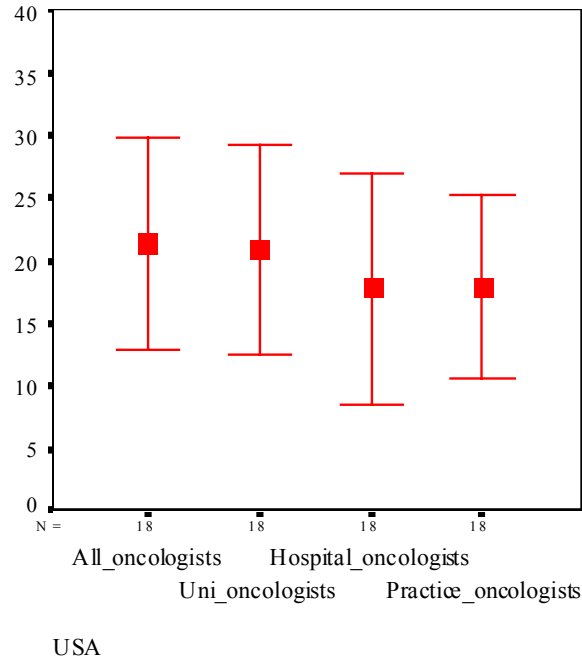


Figure 3.4: Disagreement over all 18 cases (9 per version) of the US oncologist sample (mean: 95% CI) for all oncologists (All_oncologists) as well as separated between oncologists coming from university clinics (Uni_oncologists), hospitals (Hospital_oncologists), and private practices (Practice_oncologists).

The disagreement between the three German groups differed to a small sized nonsignificant extent ($F(2/51) = 2.07, p = .137, \omega = .19$). For the US sample, quite the same was found for the three groups' disagreement ($F(2/53) = .21, p = .812, \omega = .01$), while for this sample the differences showed almost no effect. With respect to the disagreement of these three groups between countries, neither disagreement within the group of university oncologists ($t(34) = .83, p = .411, r = .14$) nor those within oncologists coming from hospitals ($t(34) = .85, p = .336, r = .14$) or from private practices ($t(34) = -.558, p = .581, r = .10$) showed to be significantly different from each other.

Happy fitting?: Description of oncologists' test decision policy. In the main study section, I hypothesized that a fast and frugal model will provide a comparable good fit to oncologists' test decision-making data as a compensatory integration model does (H3a). Due to limitations

of the design, I opted for Franklin's rule and Dawes' rule as representatives for a compensatory integration model, instead of commonly used regression models, and for the Matching Heuristic as a fast and frugal model.

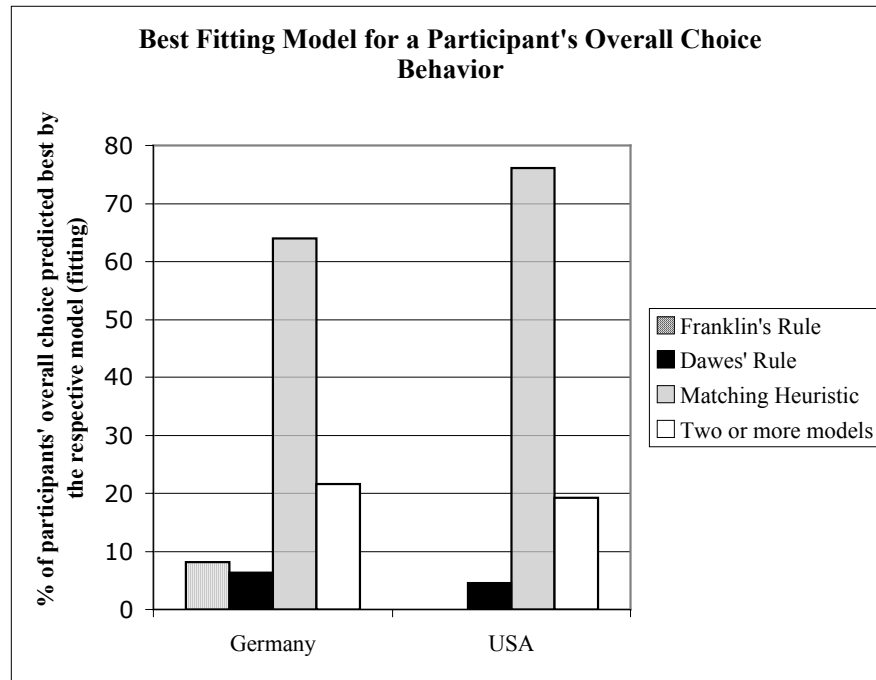


Figure 3.5: Proportion of the Best Fitting Model for a Participant's overall Choice Behavior provided by either Franklin's rule, Dawes' rule, Matching Heuristic, or by two or all Models for the German and the US oncologists.

With respect to which model out of the three applied fitted a single participant's choices best, that is, was better than the other two models at correctly predicting a participant's overall choices, Franklin's rule provided the best fit for 8.11% of the German, but none of the US participants, Dawes' rule fitted 6.31% of the German oncologists' choices and 4.59% of those of the US oncologists best, while the Matching Heuristic proved the best fit for 63.96% of the German choices and an outstanding 76.15% of the US ones. For the remaining 21.62% of the German and 19.26% of the US participants, either two or all three models provided an equally high fit for their choices. Note, that these findings do not necessarily have an implication on the general achievement of a model to fit data well, on average. A model can have an achievement of 100% for a participant, that is, it fits all choices correctly for a participant, but shares this overall fit with another model. Thus, the previous results give information about how sensitive a model is to fit a single participant's choice, on

average, better than the other two models, that is, to share least best overall fits with other models. Results are shown in Figure 3.5. Investigating any intragroup difference with respect to which model fitted the data best, calculating a Chi-square test was not possible within, and between, the two samples, as the sample size for Franklin's rule as well as for Dawes' rule was too small and led to expected frequencies in the respective cells of constantly lower than 5.0, which violated a major assumption of the Chi-square test procedure. When putting those participants together whose data were best fitted by either Franklin's rule or Dawes' rule in order to increase the sample size, and then comparing these participants with those best fitted by the Matching Heuristic for any intragroup differences, again for most of the variable Chi-squares, assumptions were violated (age, attendance of meetings, attendance of training, time for working through questionnaire; *for the US sample in addition*: location, position), for those where it was not (location, position—*for German sample only*; research involvement, years of experience) no significant differences were found.

Over all participants, the average fit of the Matching Heuristic model was 83.28% for the German sample (range = 33%–100%) and 84.20% for the US sample (range = 56%–100%). For the Dawes' rule, the average fit of the model was 58.36% (range = 0%–100%) for German oncologists and 54.13% (range = 22%–89%) for their US counterparts. Furthermore, for the Franklin's rule, here the average fit for the German sample was 67.57% (range = 33%–100%) and for the American sample 57.39% (range: 33%–89%). Figures 3.6 and 3.7, respectively, exhibit these results. Within the German sample, differences in the degree of fit over individuals indicated, between each of the three models a significant, although small sized difference ($F(2/220) = 88.66, p = .000, \omega^2 = .24$). For the US sample, identical results were found ($F(2/216) = 178.72, p = .000, \omega^2 = .14$) with respect to the overall three model comparison, however, here Franklin's rule and Dawes' rule did not differ significantly from each other. Between both countries, this difference in range was significant and medium sized for Franklin's rule ($t(211.25) = 5.05, p = .000; r = .34$) as well as for Dawes' rule ($t(218) = 2.11, p = .036; r = .17$) where rather a small effect was found, while this was not the case for the Matching Heuristic ($t(218) = -.416, p = .776; r = .02$).

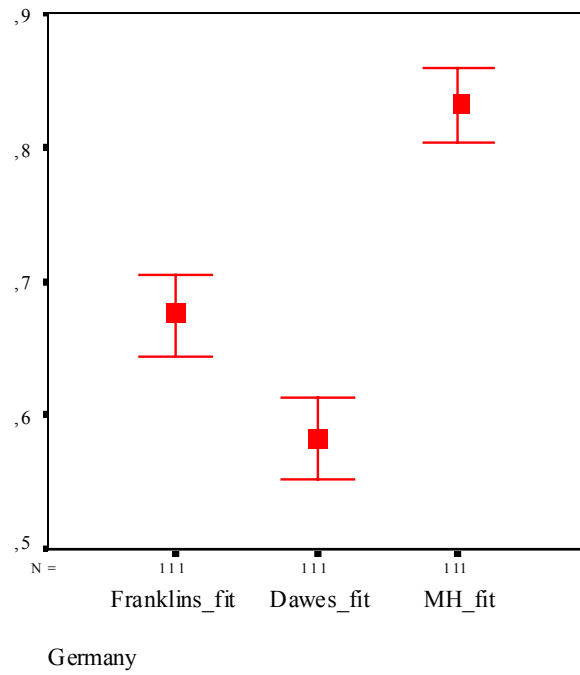


Figure 3.6: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German oncologists ($N = 111$).

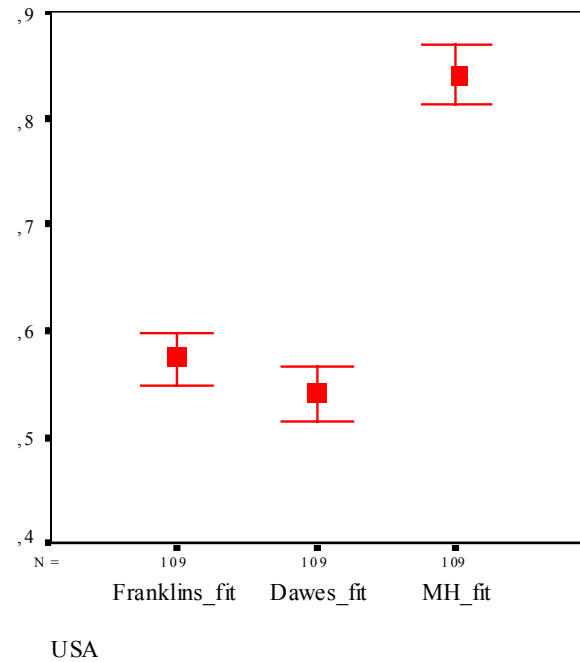


Figure 3.7: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US oncologists ($N = 109$).

Given these findings, I consider the Hypothesis 3a as being confirmed in that the Matching Heuristic model provided an equally good fit, and in this study even a better one, to oncologists' test decision-making data than both of the compensatory integration models used here.

How general are the models?: Generalization performance of the three models. A model's superiority in terms of being a useful representation of reality, however, is commonly measured by its generalization performance than by its fitting. We, therefore, bravely hypothesized, additionally, that a fast and frugal model will also provide an equally good prediction to unknown data (generalization) as a compensatory integration model does (H3b).

When looking at the models' generalization performance, the Matching Heuristic's lead vanished, however. Across all participants, Franklin's rule provided the best prediction of unknown cases, that is, counted most hits for a single participant, compared to the other two models,²⁶ for 53.15% of the German and for 46.79% of the US participants. The Matching Heuristic performed better than the other two models for 36.94% of the German participants, while it did still for 44.03% of the Americans. Dawes' rule was able to do so for 3.60% of the German participants and for 1.83% of the American ones. For the remaining 6.31% of the German participants and for 7.35% of the US ones, either two or all of the three models provided an equally overall prediction. Figure 3.8 exhibits the data. Regarding the investigation of any intragroup differences, I found exactly the same situation as already reported in the "fitting" sections.

Across all participants, Franklin's rule provided the best average prediction of unknown cases for 68.07% (range = 33%–100%) of the German participants and for 57.70% (range = 33%–89%) of the US ones. The Matching Heuristic's average with respect to this was 55.16% (range = 11%–100%) for the German sample and 51.48% (range = 0%–100%) for the US one, while Dawes' rule provided an average prediction of 35.98% (range = 0%–89%) for the German sample and for the US sample of 41.54% (range = 0%–89%). Figures 3.9 and 3.10, respectively, demonstrate the results per country. Within the German sample, differences in the degree of the prediction of unknown cases over individuals indicated a significant,

²⁶ Note that low proportions do not necessarily imply that the respective model has a low prediction performance, on average. Instead, results refer to a model's ability to predict the overall choice for a single participant better than the other two models, and in that it does not share an equal performance with another model.

almost medium sized difference between the models ($F(2/220) = 58.77, p = .000, \omega^2 = .29$). For the US sample, I found an almost identical result ($F(2/216) = 52.92, p = .000, \omega^2 = .14$), however, differences here were found to be small sized. Between both countries, the models' generalization ability were significantly different, but small for Franklin's rule ($t(211.82) = 5.18, p = .000, r = .11$) and for Dawes' rule ($t(217.92) = 2.09, p = .038; r = .14$), while it was not for the Matching Heuristic ($t(218) = -.953, p = .342; r = .06$).

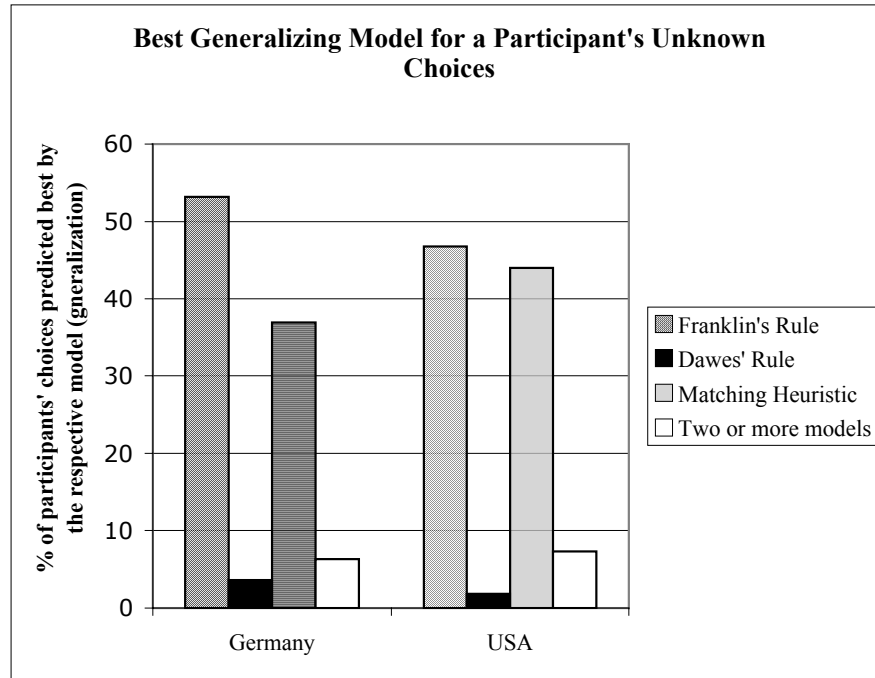


Figure 3.8: Proportion of the best generalizing model for a participant's overall choice behavior provided by either Franklin's rule, Dawes' rule, Matching Heuristic, or by two or all models for the German and the US oncologists.

With respect to the generalization setting, the Matching Heuristic still performed fairly well given its parsimony in cue use, and even outperformed one of the compensatory competitors, namely, Dawes' rule. Nevertheless, on the basis of the findings for both countries, I cannot confirm the Hypothesis 3b that it predicts the unknown choices of oncologists for a task, such as presented equally, well as a compensatory integrative model would. Here, Franklin's rule outperformed the Matching Heuristic and made a better prediction for unknown cases.

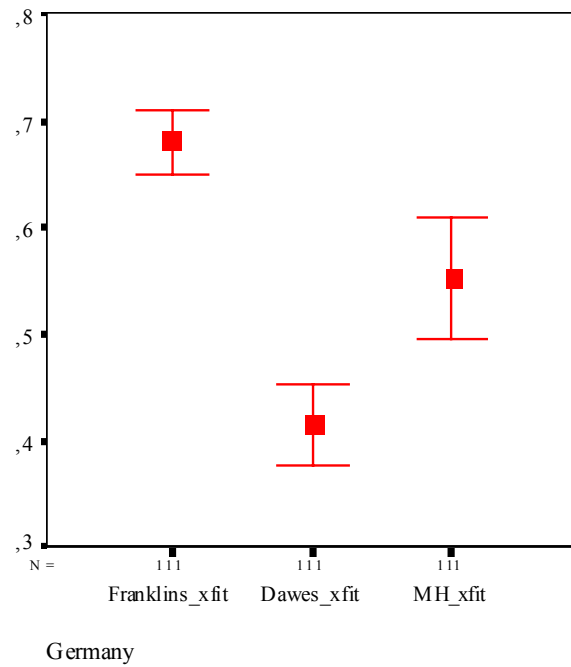


Figure 3.9: Results of average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German oncologists ($N = 111$).

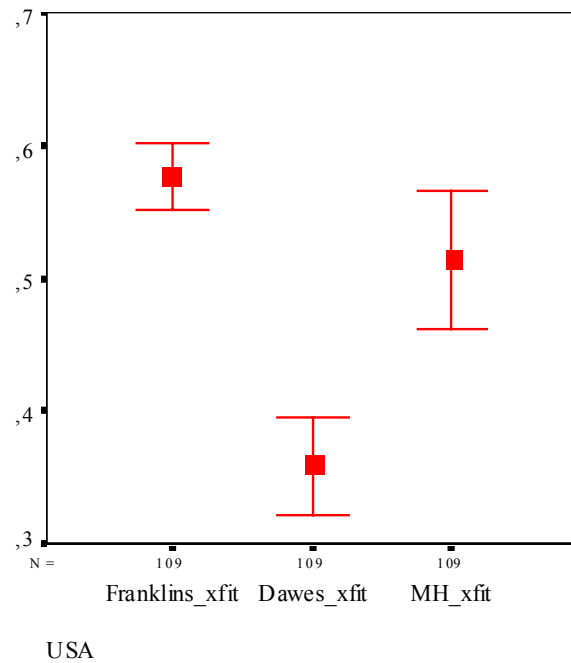


Figure 3.10: Results of average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US oncologists ($N = 109$).

Taking all or one?: Oncologists' cue use. Related to the question of which model best predicts data is the issue in which cues appeared to be important. While I found Dawes' rule to be of

minor importance in explaining both of the oncologists' choices, Franklin's rule showed to be best in generalizing oncologists' data, while the Matching Heuristic performed best in their overall fitting and fairly well in the prediction setting, given its parsimony. For this reason, I will give, for both Franklin's rule as well as the Matching Heuristic, an impression of what cues these models used in order to make their predictions.

To start with the Matching Heuristic for the fitting situation, cue use is defined broadly as the number of cues searched including the cue on which the final decision is based, whereby the number per case may vary. The number of cues used over all nine cases was calculated for each oncologist. Across all oncologists, the mean number of cues used ranged from one to three (*mean*: 1.67; *SD*: .839), while it ranged also from one to three (*mean*: 1.46; *SD*: .757) for the German sample, it also did for the US sample (*mean*: 1.67; *SD*: .839). Although not apparent at first glance, when comparing the number of cues used between both countries, a significant, small effect sized difference between the two samples was found ($\chi^2(2) = 139.57, p = .000, \phi = .27$). Taking a closer look at the number of cues used, it was found that, within the German sample, 70.30% of all choices (N=999) were explained by a K=1 model, but only 13.50% by a K=2 model, and 16.20% by a K=3 model. For the US sample, however, only 44% of the choices (N=981) were explained by a K=1 model, while still 23.9% were by a K=2 model, and 32.10% by a K=3 one.

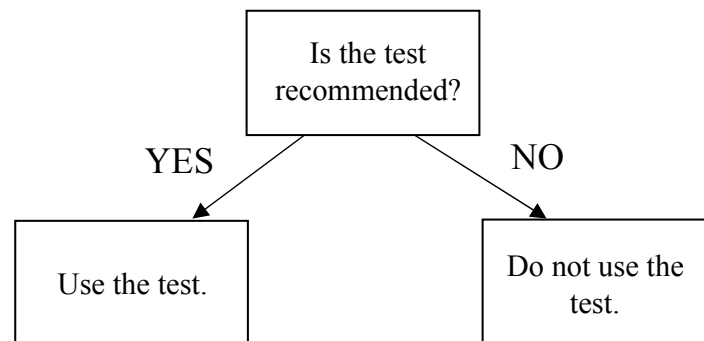


Figure 3.11: Example for a German K=1 model (70.30% of all German choices) displaying the most frequently used cue.

Comparing the Matching Heuristic across oncologists of each country, it was found that they differ in terms of the cues they used to make their decisions on ordering the test. For the German oncologists, 54.05% of the participants made use of the information about

whether the test was recommended by guidelines. The information about the degree of undertreatment was used by 36.04% of them, followed by information about the test's costs (35.94%), and by information about the side effects related to the treatment (20.72%). In contrast to this, the US participants mostly used the information about the test's cost (59.63%), closely followed by guideline recommendation (57.49%), and by information about the risk of undertreatment accompanying the test (54.33%), while only 17.43% used the information about the treatment's side effects. Figure 3.11 provides an example for a German oncologist employing the most frequently occurring K=1 model and the most frequently used cue.

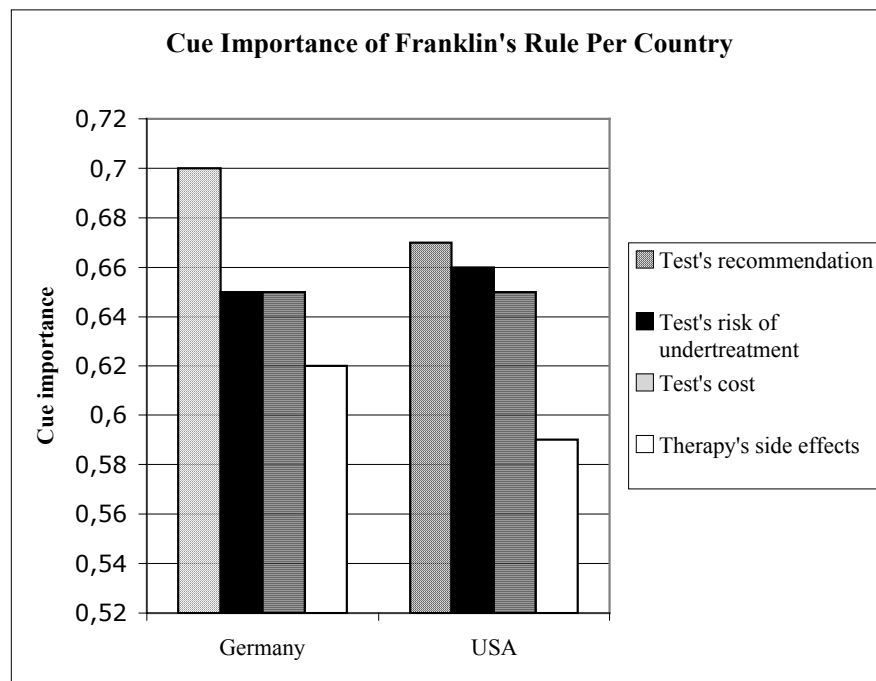


Figure 3.12: Cue importance of Franklin's rule for the German as well as the US oncologist sample in the generalization setting.

For Franklin's rule, it was outlined that this model assumes that people take all cues into account, and to put corresponding weights of importance on each of them. In accordance with the findings of the Matching Heuristic for the German sample, also in the case of generalization it was found by Franklin's rule for this sample again that recommendation by the guidelines had the highest relevance for participants (.70), followed by the undertreatment cue (.65), the cost of the test cue (.65), and the cue referring to the side effects of the treatment (.62). Within the American sample, results proved to be different to the findings of the

German sample: Here the undertreatment cue received the highest weight (.67), information about a test's cost took the second position with a weight of .66, while information about a test's recommendation by the guidelines (.65) and side effects of the treatment proved to have the smaller weight (.59). Results for both countries are shown in Figure 3.12.

Reported representativeness of the questionnaire. With respect to the felt representativeness of the questionnaire, the majority of both the German as well as the American oncologists' sample regarded answering the questionnaire as easy and the described situation as straightforward. Moreover, the decision criteria generally offered in the scenarios were found to also be the criteria they would use in such a medical decision situation, which leads to the fact that only a minority of each sample requested further information (see next paragraph—requests for further information). The time participants needed to work through the scenarios ranged for the German sample from 1 to 30 minutes (mean: 8.94; *SD*: 4.12) and for the US participants from 2 to 20 minutes (mean: 8.61; *SD*: 3.70). Investigating for differences between the two samples with a Chi-square test for the variables F1 to F3, and with an independent t-test for variable F5, showed only small and nonsignificant differences for each of the tested variables. Table 3.8 exhibits all respected values in detail.

Table 3.8: *Values of Reported Representativeness of the Questionnaire*

<i>Question</i>	<i>Germany (N = 111)</i>	<i>USA (N = 109)</i>
F1	(1) N = 85–76.6% (2) N = 21–18.9% (3) N = 5–4.5%	(1) N = 92–84.4% (2) N = 12–11% (3) N = 5–4.6%
F2	(1) N = 88–79.3% (2) N = 23–20.7% (3) N = 0	(1) N = 81–74.3% (2) N = 28–25.7% (3) N = 0
F3	(1) N = 94–84.7% (2) N = 17–15.3% (3) N = 0	(1) N = 79–72.5% (2) N = 29–26.6% (3) N = 1–0.9%
F4	N = 17–15.3% See below (requested information)	N = 30–27.5% See below (requested information)
F5	1–30 minutes (mean: 8.94; <i>SD</i> : 4.12)	2–20 minutes (mean: 8.61; <i>SD</i> : 3.70)

Requests for further information. A total of 17 requests for more information within the German (N = 111) and a total of 30 within the US sample (N = 109) were made by oncologists when asked if they missed information within the scenarios that they would normally use for their medical decisions (F4). For the German sample, 29% of these 17 oncologists requested

information about patient's preferences on having such a test. Eighteen percent of the German requests were for information regarding the nature or rate, respectively, of the therapy's side effects, 6% asked for the Karnofsky index, and 12% for the patient's prognosis. Thirty-five percent indicated to have missed information within the scenario, but did not make any suggestions which exactly ones.

For the US oncologists, 53% of the 30 oncologists requesting further information were interested to know the patient's preferences in having such a test, 33% wanted to know more about the nature or rate, respectively, of the therapy's side effects, 10% requested more details on either the prevalence or incidence of the disease, and 3% wanted to know about the basis of evidence.

Findings beyond fitting and predicting. Within the pilot study section, I proposed some further hypotheses beyond fitting and predicting. In the following, the results for these hypotheses will be presented.

Based on the findings of the pilot study, I assumed that whenever a pharmacodiagnostic test is recommended by guidelines and protocols, respectively, oncologists choose to use the test regardless of the other cue values. This idea was formulated in Hypothesis 1a. Taking a glance at the cases²⁷ where the recommendation cue received the value for a recommendation by way of the guidelines, it was found that the German sample that opted for the test in those cases ranged from 95% to 98% (*mean*: 96.33; *SD*: 1.37), which was almost identical to the US sample where opting for the test ranged from 96% to 98% (*mean*: 97.33; *SD*: 1.03). For the cases²⁸ where the recommendation cue received the value for a nonrecommendation by way of the guidelines and experts, I observed for the German sample that, here, opting for the test ranged from 16% to 55% (*mean*: 36.83; *SD*: 13.79), while for the US sample it ranged from 24% to 52% (*mean*: 40.33; *SD*: 11.67). When comparing these cases with the respective recommendation value with these case with the nonrecommendation value within the German sample, a large sized significant difference in the proportion for deciding to do the test in favor of these cases was found, where doing the test was recommended ($t(110) = 11.88; p = .000, r = .75$). The same was observed for the US sample

²⁷ These cases were: 3A, 7A, as well as 8A, within the A version of the questionnaire, and 2B, 3B, and 7B for the B version.

²⁸ These cases were: 4A, 5A, as well as 9A, within the A version of the questionnaire, and 4B, 8B, and 9B for the B version.

($t(108) = 11.76; p = .000, r = .74$). Across both countries I found no differences with respect to the proportion for opting for the test when the case had the critical recommendation value ($t(218) = -.84; p = .403, r = .06$), as for these cases doing the test was not recommended by the respective guidelines ($t(218) = -.69; p = .490, r = .05$). Proportion data are shown in Figure 3.13.

Given these results, I confirm Hypothesis 1a for both countries, that oncologists decide more on conducting recommended tests than they would for nonrecommended ones.

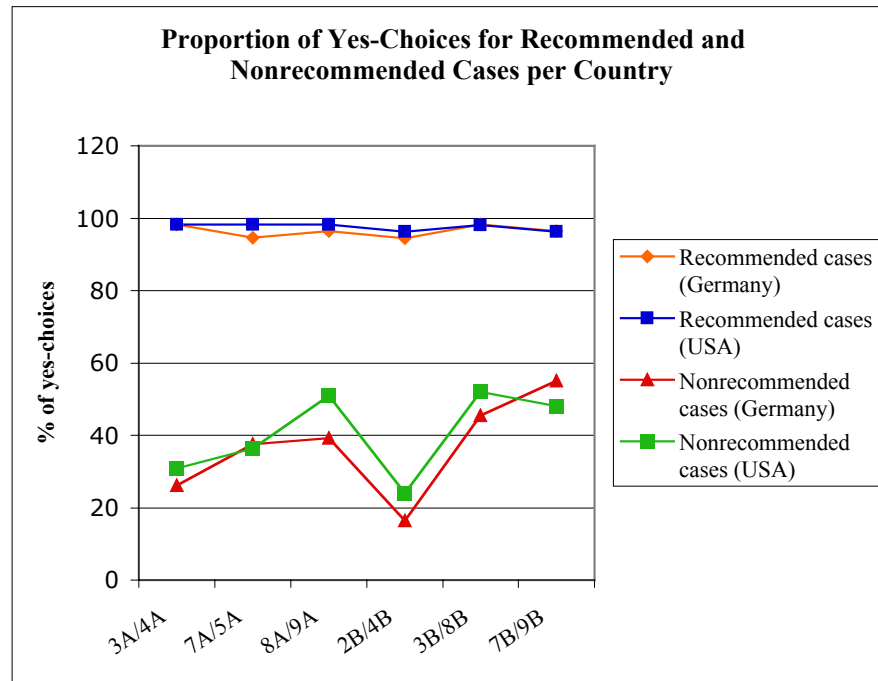


Figure 3.13: Proportion of yes-choices (case wise) for recommended and nonrecommended cases per country (H1a).

It was furthermore hypothesized, that, in case the severity of the side effects of the treatment increases, it will be found more choices for tests, which do not have a recommendation of guidelines (H1b). In order to test this hypothesis, I compared the proportion of yes-choices for the cases having cue profiles of less severe side effects (serious side effects during the treatment) and a nonrecommendation²⁹ with the cases where the cue profile includes the more severe side effect cue value (permanent serious side effects) and again the nonrecommendation cue value.³⁰ For the German sample, the proportion for opting

²⁹ These cases were: 4A, 5A, as well as 4B.

³⁰ These cases were: 9A, 8B, as well as 9B.

for a test if it was without a recommendation and for a treatment with less severe side effects, ranged from 16% to 38% (*mean*: 27.00; *SD*: 11.00), while for cases in which the side effects for the treatment were described as more severe, this ranged from 39% to 55% (*mean*: 46.67; *SD*: 8.02). That is, German participants' proportion to opt for a nonrecommended test, which was linked to a more severe treatment, was found to be significantly higher than that for such tests which would make statements for less severe treatments ($t(110) = -3.24$, $p = .001$, $r = .30$). The difference of choice behavior in this regard showed again a medium sized effect.

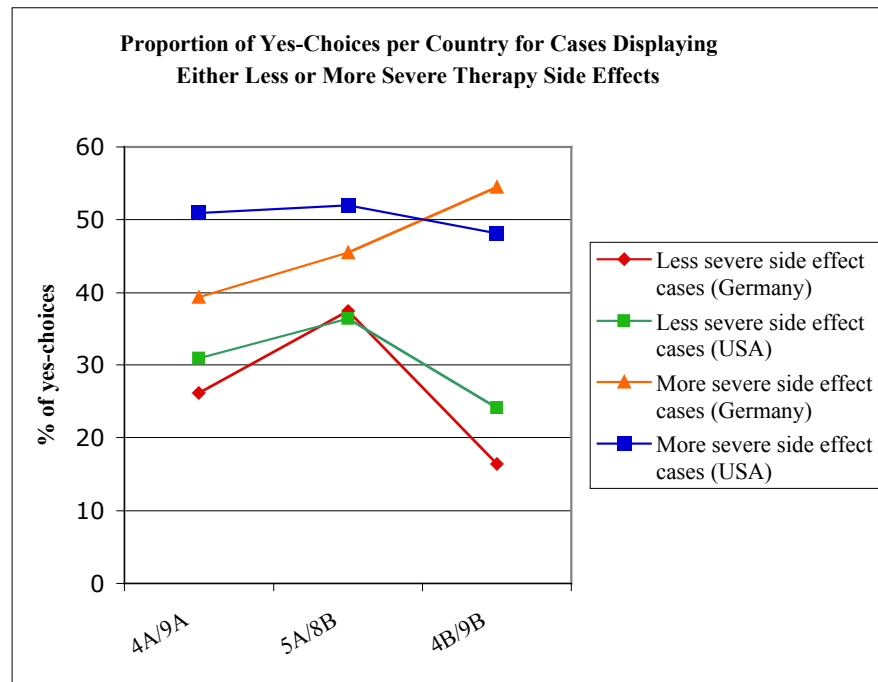


Figure 3.14: Proportion of yes-choices (case wise) per country for nonrecommended cases either having the less or the high side effect cue value (H1b).

In accordance with these findings, US participants also decided more often to do a nonrecommended test when it was linked to a treatment with more severe side effects (range: 48%–52%; *mean*: 50.33; *SD*: 2.08) than they did for the less severe treatments (range: 24%–36%; *mean*: 30.33; *SD*: 6.03). This difference in proportion proved to be medium sized and also significant ($t(108) = -3.31$; $p = .001$, $r = .30$). Comparing the yes-choice proportion for both conditions across countries, neither where side effects were less ($t(218) = -.542$; $p = .588$, $r = .04$) nor where side effects were high ($t(218) = -.550$; $p = .583$, $r = .04$) was a meaningful difference observed. Proportion data for both countries are shown in Figure 3.14.

In the light of these findings, which proved a significantly higher proportion of yes-choices for a nonrecommended test that is connected to a treatment with more severe side effects than for one that is not, I consider the Hypothesis 1b to be also confirmed.

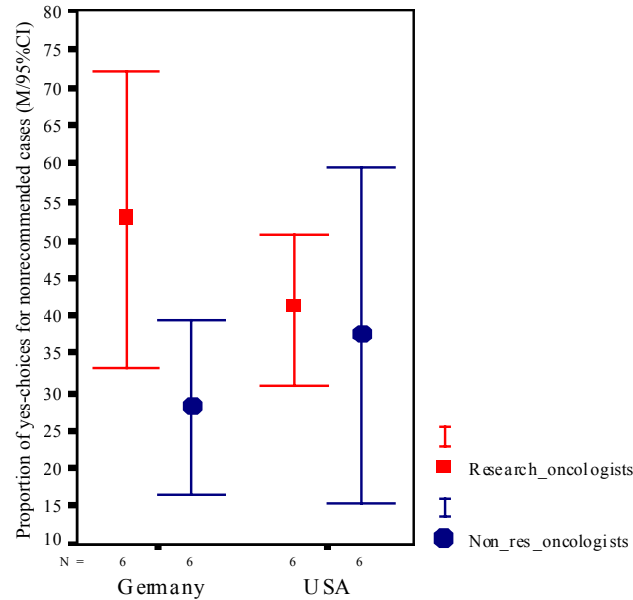


Figure 3.15: Proportion of the yes-choices (mean: 95% CI) per country of research-involved oncologists (Research_oncologists) vs. nonresearch involved oncologists (Non_res_oncologists) for nonrecommended cases (three cases per questionnaire version).

Another hypothesis developed on the basis of the findings of the oncologists' pilot study stated that a recommendation by guidelines is less needed for deciding on having a test by oncologists who were involved in clinical research than it would be by participants who were not involved. If this were true, I should find, for the oncologists who are involved in research, that the relative proportion of yes-choices on cases which did not have the "recommended by guidelines" cue value should be higher than that for those oncologists who are not involved. For German research involved oncologists, I found relative proportions of yes-choices for cases having the nonrecommendation cue value, which ranged from 19% to 69% (mean: 52.62; SD : 18.71), while for nonresearch oncologists it ranged from 14% to 46% (mean: 27.98; SD : 11.04). This difference between the two groups, which showed almost a medium effect size, proved to be significant in favor of the research involved oncologist ($t(108) = 3.21, p = .001, r = .29$). For those oncologists of the US sample who were involved in

research, relative proportions of yes-choices on not recommended tests ranged from 24% to 50% (mean: 40.94; *SD*: 9.43), and from 15% to 60% (mean: 37.44; *SD*: 21.04) for those participants not involved in research. For the US sample, I saw no difference between the two groups, however ($t(107) = .20$; $p = .423$, $r = .02$). That is, while indeed German research oncologists did decide more often to order a nonrecommended test than the nonresearch involved counterparts, this was not found within the US sample—I, therefore, regard Hypothesis 2a only to be confirmed for the German sample.

Regarding any intercultural differences, neither for the two groups of nonresearch oncologists' ($t(94) = -1.38$, $p = .171$, $r = .14$) nor for these of the research oncologists ($t(121) = 1.32$, $p = .189$, $r = .12$) a meaningful difference was found. Figure 3.15 exhibits the mean and the 95% confident interval for the proportion of yes-choices for nonrecommended cases per country.

Within the last remaining Hypothesis 2b, I proposed that research involved oncologists would show no significant differences regarding the applied decision-making structure, compared to their nonscientific counterparts. While it was mentioned already that, in the case of the best fitting model, the sample sizes for Franklin's rule and Dawes' rule were too small for making any meaningful intragroup comparison in the case of determining the best predicting model, I found quite comparable large groups for Franklin's rule and the Matching Heuristic. I chose all the participants best predicted by either of these two models and investigated if I could find the number of research and nonresearch oncologists within one respective model to be different from the other. The answer was no. Within this setting, German research oncologists' choice behavior was best predicted to be 56% for Franklin's rule and 37% for the Matching Heuristic, compared to the nonresearch oncologists in this sample, for which Franklin's rule predicted 52% of the unknown cases best, and the Matching Heuristic 36%. Across the two best predicting models, no significant difference was found in the number of research or nonresearch oncologists ($\chi^2(1) = .005$, $p = .946$, $\phi = .01$). The US sample did not show any other results. Here, research oncologists' choice behavior was best predicted to be 45% for Franklin's rule and 47% for the Matching Heuristic, while for the nonresearch oncologists Franklin's rule predicted 52% and the Matching Heuristic 36%. When comparing the number of research or nonresearch oncologists across the two best predicting models identically, I did not find a significant difference ($\chi^2(1) = .829$, $p = .363$, $\phi = .10$). Based on these results, I can conclude that research oncologists and nonresearch oncologists

do not differ in terms of applied decision-making strategies, which confirms the final Hypothesis 2b.

3.6 Summary & Discussion

The aim of the main study was to learn about oncologists' decision-making strategies and cue use when faced with a decision on a pharmacodiagnostic test for tailoring a cancer treatment decision. Data were derived from nine presented hypothetical cases, which were developed on the basis of findings of an upfront pilot study. In order to answer the questions of the main study I confronted the data with two compensatory, integrative models, namely Franklin's rule and Dawes' rule, and one noncompensatory model, which was the Matching Heuristic. I investigated the relative ability of these three decision making models to describe and predict decisions on a pharmacodiagnostic test made by German and US oncologists in response to systematically varied case vignettes.

It was found that both German and US oncologists were quite ready to apply a pharmacodiagnostic test to their treatment decisions, while US oncologists were even more prepared to do so than their German counterparts. Nevertheless, it was also seen that oncologists, regardless of what country or location they were from, disagreed to some extent with regard to the decision to be made on the same case. Additionally, German as well as US oncologists' test decisions were better described by a fast and frugal model called the Matching Heuristic than by either of the two compensatory integrative models, namely Franklin's rule and Dawes' rule, while Franklin's rule made better predictions to unknown cases than either the Matching Heuristic or Dawes' rule did. It was also seen for both countries that a recommendation of the test by guidelines nearly always resulted in a choice for doing the test, which differed significantly from the choice behavior for cases in which the test was not recommended. However, when the side effects of the treatment the test focused on were described as likely to increase, significantly more choices for conducting a nonrecommended test anyway were noticed for both German and US oncologists. Finally, while German oncologists who were involved in research were influenced less by guidelines in that they opted significantly more often for nonrecommended tests than their nonresearch counterparts did, such a difference could not be found for the US sample. However, for both samples it proved true that neither German research-involved oncologists nor US ones differed from their nonresearch colleagues with respect to the applied decision-making strategy.

Oncologists' rather high willingness to opt for a pharmacodiagnostic test certainly reflects what I already referred to earlier in this Chapter – the desire of clinicians in the field of oncology to utilize as many beneficial options as possible of the few at hand. However, it may furthermore mirror oncologists' desire to get a tool at hand that would bring more rationale in a field where not much consensus exists on standard care and thereby to reduce at least some uncertainty that accompanies their daily treatment decisions.

This uncertainty about the best course of action to take may be also reflected in the degree of disagreement between oncologists found in this study. Such variations in judgment have been also elicited for other groups of clinicians such as general practitioners by various studies (e.g., Gillis et al., 1981; Evans et al., 1995; Dhami & Harries, 2001). However, the investigated groups were usually not faced with life-altering medical decisions. That I arrived at comparable findings may just outline also for oncologists what was found for other kind of clinicians as well, that is, that they do not seem to have a common formula such as Bayes' theorem (see Chapter II) at hand with which they systematically assess when e.g., to apply a test would be appropriate and when not. If they had they should make less disagreeing choices. This is not to suggest that there is a perfect solution to any of the presented cases, which is regarded as being the accurate one by the investigator. It is simply to highlight that seemingly the proposed ideas within medical textbooks for helping clinicians in deciding when to test might be not well accepted or helpful at all. I can appreciate that in tasks such as those presented in this study there is a high degree of uncertainty according to what each single piece of information indicates for the outcome, as it is for most situations in medicine and probably in most other real-life settings. This certainly makes the variations in choice more understandable, while not always desirable.

Perhaps not surprisingly for a person who is in charge of making fundamental and highly impacting decisions on life and survival, I found that the most complex, compensatory integrative model applied to the data in this study, namely Franklin's rule, was better able than the other two models in predicting oncologists' test decision policy. This finding receives much support from numerous studies showing that people confronted with decision problems of a high significance and importance, respectively, and with outcomes not reversible, search for more information and apply more complex decision-making strategies (e.g., Billings & Scherer, 1988; Ford, Schmitt, Schechtman, Hults, & Doherty, 1989; McAllister, Mitchell, & Beach, 1979; Christensen-Szalanski, 1980). However, since the medical decision-making

setting is characterized by task complexity and time pressure (see Chapter II), one could have expected that individuals tend to simplify the cognitive requirements of the decision process because they are limited in their information-processing capacity. But then, it should have been found that their decision-making strategies are equally well predicted by a fast and frugal model. I assume there to be some further reasons other than the significance of the test decision why this hypothesis could not be confirmed. In contrast to oncologists' daily workload, which includes time pressure for certain, there were no time constraints for this test situation. However, if time constraints for solving the decision task are lenient or not really existent, more complex decision strategies are significantly more likely to be applied (e.g. Abelson & Levi, 1985; Payne et al., 1988). In addition, I was not able to manipulate as many cues as I had found to be important for the choice on a pharmacodiagnostic test during the pilot study. The limits were set by the focused study population. Presenting them with more than 10 cases would have remarkably decreased the likelihood of convincing a sufficient amount of participants to participate in the study when not offering any compensation. For that reason and to ensure an orthogonal cue design this left me with no more than four cues to manipulate, which surely reflects only to a certain extent of information load oncologists face within their daily practice. Nevertheless, 36.9% of the German participants' choices and 44.0% of the US ones were still better predicted by the noncompensatory Matching Heuristic than by either of the two compensatory models. However, while both of the compensatory models made use of all four cues for the prediction, the Matching Heuristic model searched an average of 1.68 cues (averaged across both samples). Given the parsimony of the Matching Heuristic in cue use it still performed fairly well in predicting the oncologists' test decision-making policy and even outperformed the other compensatory competitor, Dawes' rule.

That the use of heuristics can be far away from fallibilities and biases may even convince authors who have pictured these in such a light, e.g. Chapman and Elstein (2000) or Kahneman and Tversky (1982), when taking a closer look on those heuristics applied in this study. Over 50 percent of all of participants' choices were predicted by a Matching Heuristic model that was based on the single information about whether the test is recommended by guidelines or not. Although there was no accuracy criterion in this study for when the application of the test would be appropriate, especially in a field such as oncology where decisions have an tremendous impact on a patient's survival and where the preferable standard care may be less well definable than for other medical fields, it might be rather reasonable to

rely on guidelines³¹. Despite all disagreement, in cases in which the test was recommended an average of 96.8% of the participants (across both samples) opted for doing the test in such situation. It may be discussable to what extent these answers were given socially desirable. Within the pilot study, however, following guidelines did not appear to be regarded as a kind of decision making to show off with by oncologists, while still a proper one. Instead they seemed to spend more effort on ascertaining the investigator to what a high extent their medical decisions are still made up by their own knowledge and experience. In addition, physicians are not obliged to make use of them. For that reason I believe that these findings are more likely to reflect the reality than a picture of pretending to behave in a socially desirable way. This assumption may also be supported by the finding that as soon as side effects of the treatment increased, oncologists opted significantly more often for also applying a nonrecommended test, which they would not have done if they were interested in pretending to behave in a socially desirable way.

This observed tendency of oncologists to use nonrecommended tests, however, is viewed with some concern. Although the design of the main study did not enable me to make any clear statement about whether or not oncologists are able to understand and interpret probabilistic information properly, pilot study results indicated that they are not. That is, while a recommendation by guidelines might at least assure that no test becomes routine, which would entail even more harm to patients than current standard procedure, the same could not be stated if an oncologist were in charge of it who is hardly able to work with terms such as sensitivity and specificity. The concerns were further supported by seeing German oncologists involved in research deciding more often on also having a nonrecommended test than their counterparts not involved in research.

Some relief from these concerns comes from the result that German as well as US oncologists used information about the risk of undertreatment rather frequently for their decisions, and in that this information proved to be quite important for their choices on pharmacodiagnostic tests. Depending on whether it was looked up by Franklin's rule or the Matching heuristic, the cue usually ranked on the second and third place of importance. This is reassuring in that oncologists are at least aware of the significance of this issue. However, it is to stress that such kind of information is not always available to them.

³¹ Some problems related to guidelines are discussed in Chapter IV in the respective discussion section.

An oncologist's test decision can have significant consequences on cancer patients' lives. Even though the risk of getting false results is an inherent characteristic of each test, in the case of pharmacodiagnostic test a false result leading to an undertreatment can almost certainly mean a life-sentence to the patient. Given the fact that information about the risk of undertreatment is hardly available to oncologists, or only available in a probabilistic, incomprehensible format (see pilot study), there are serious interventions to be considered before more of such tests will hit the market. Even more so as it was found that guidelines are not always the basis for the decisions made. One point to start with would surely be to find a more comprehensible way in educating oncologists (and clinicians in general) in interpreting and understanding probabilistic information than done so far with the counterintuitive Bayes' theorem. Gigerenzer and Hoffrage (1995) suggested a simple reproducible *natural frequency tree* to resolve the task of understanding and calculating a test's predictive values (see Figure 3.16). For instance, at a medical school considering such a format the young Mr. Plaster, introduced in Chapter II, would in a lesson about how to interpret test results quite certainly be shown the following example instead of the Bayes formula (see Chapter II: *Formula 2*) given again the same values used there, that is, an assumed prevalence of the tropical disease of 30% and a test's sensitivity and specificity of 80%:

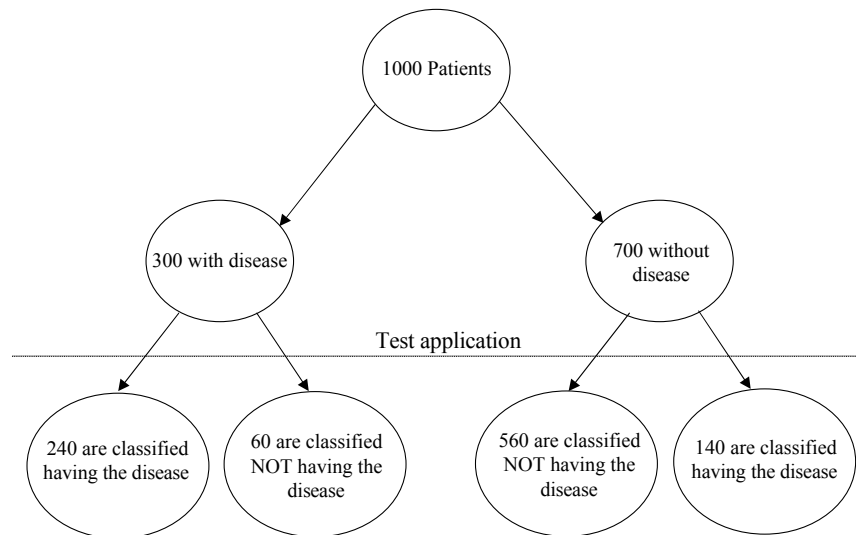


Figure 3.16: *Natural frequency tree for a hypothetical tropical disease having a prevalence of 30 percent and a hypothetical test with a sensitivity and specificity of 80 percent*

From there, he could easily find out for instance, what the likelihood of having the disease given a positive test result is, by simply taking the proportion of the true positives (240) out of all positives (240 plus 140) and come up with the same result (63%) as by using complex Bayes' theorem. Several studies done with physicians (e.g., Gigerenzer & Hoffrage, 1995; Sedlmeier & Gigerenzer, 2001) and with patients (Kurzenhäuser, 2003; Slaytor & Ward, 1998) have shown quite satisfying evidence for the advantage in better understanding probabilistic information if it is expressed in natural frequencies compared to those expressed in conditional probabilities as by Bayes theorem. Supported with knowledge about how to use such kind of natural frequency tree it might be easier for oncologists to understand and properly estimate with even only probabilistic test information at hand what risk of undertreatment the test might entail for a specific patient population. But, although human beings' conditional probability blindness is to a certain extent a good contributor, confusion about conditional probabilities is not the entire explanation for why a physician's probability estimates and integration in diagnosis and treatment are not in accord with those dictated by Bayes' theorem or other normative approaches. Environmental constraints I have already exhaustively pointed towards, such as limited time, the availability of incomprehensive information as well as the general limited computational capacity of human beings might make it in general rather unlikely to always handle the array of collected data sufficiently.

Therefore, guidelines could be one, if not the adequate decision aid to help oncologists to make proper decision in accord with what has currently proved to be the gold standard in diagnosing and treating diseases. However, doctors seem to not always regard them as a tool worth leading their medical choices. The reasons for that might be various. It might partly reflect clinicians' overconfidence in their own judgmental abilities, as it appeared in the pilot study. Other findings (e.g., Ustun & Sartorius, 1995; Smith et al., 2003) point towards the issue of some guidelines' complexity, while within the pilot study some participants regarded guidelines to be not sufficiently up-dated. That is, there are seemingly some open tasks to solve for those in charge of developing medical guidelines. In a study conducted on the decision of when to send a patient into the coronary care unit instead into the nursing bed, it was found that a simple decision tree consisting only of three simple questions were more helpful and led to equally accurate decisions than a decision aid applying more than fifty pieces of information (Forster et al., 2002). Less is often more also in the field of medical decisions. A finding such as this should definitely encourage people in charge of setting up

guidelines to develop them in a manner that is consistent with simple models that appear to be used naturally by human beings. For instance, to outline guidelines by using a simple decision tree might increase their acceptance. This acceptance will be needed if aspired that medical decisions meet the demands of the current best standard of care. And who knows, by not aiming at compiling long arrays of all available data within guidelines anymore, but by concentrating on the most important ones, it might make it even possible to keep guidelines updated fast *and* frugal.

In Chapter VI I will discuss methodological concerns related to the study.

4 Study 2: What Role Do The Pathologists Play?

“The art of being wise is the art of knowing what to overlook.”

William James

During the oncologists’ pilot study, to my surprise it appeared that pathologists play a role in the decision-making process of when to use a particular pharmacodiagnostic test. That is, once the oncologist has decided that testing is needed, the pathologist is involved in determining which of a number of available tests is most appropriate. Therefore, the features of a test, such as sensitivity and specificity, should be more meaningful to pathologists than found them to be for oncologists. To clarify what role pathologists play in decisions on pharmacodiagnostic tests, a separate study of pathologists was conducted.

4.1 Pilot Study

4.1.1 Design

In a pilot study identical to that used for the oncologists, semi-structured interviews with an open-answer format were conducted to examine pathologists’ decisions regarding pharmacodiagnostic tests. The aim of this pilot study was identical to that of the oncologists’ study. Consequently, also for this group the interview questions focused on how pathologists usually arrived at their decisions to perform a test in general, whether they were already experienced in using any of the target tests available, and if so, what made them start performing these tests. Further questions addressed whether the decision to perform different kinds of test would make them apply different criteria. Finally, other questions were designed to provide insight into the relationship and flow of influence between oncologists and pathologists. Table 4.1 summarizes each possible question.

4.1.2 Procedure

Apart from ensuring that pathologists were recruited from all quarters, that is, university clinics, hospitals, and private practices, participants were chosen at random. Each pathologist was contacted by phone and asked to participate in this pilot study. Telephone numbers and basic information about each of the pathologists were obtained from the internet.

When they agreed to participate, they were interviewed at their place of work. All participants were told their answers would remain completely anonymous and confidential. The interviews were tape-recorded and ranged from 30 minutes to 3 hours (Germany mean: 73 minutes; USA mean: 73 minutes), depending on how willing participants were to go beyond simple answers to the questions.

Table 4.1: Catalogue of Interview Questions for Pathologists

Topic	Questions
Demographic data of interviewee	Age Gender Qualifications Position Years of experience Work location
1. Questions on general decision making	1.1 Please explain the process of how a test moves from “available on the market” to “used in the lab”. 1.2 Are there any defined criteria that determine whether a test will be used? 1.2.1 If you have more than one test at hand for the same diagnostic question, what criteria lead you to your choice of the most appropriate one? 1.3 Some tests are recommended in guidelines. However, additional tests are commercially available. Do you take those into account? 1.3.1 If yes: What criteria determine whether such a test is used? 1.3.2. If no: Why is the decision against such a test made? 1.4 Do you remember any case in which you stopped using a test or refused completely to use it? Short explanation of predictive and prognostic marker detection 1.5 What criteria would you apply for such tests to decide to implement them in your laboratory? 1.6 Would specific features of the test play a role? 1.6.1 If so: Which ones? <i>If sensitivity, etc. are mentioned: Do you know the figures for that for the currently used tests.</i> 1.7 Are there any features of the test that would limit its use?
2. Questions concerning target tests	2.1 Have you already have had any experience with tests opportunities that have given you information about the efficacy of cancer treatments or specification for a suitable therapy? (Target tests) 2.1.1 If so: Do you use it and why? 2.1.2 Does the underlying technology of a test play a decisive role in your decision to use it? (Or is rather the fashion way in which the test is carried out relevant or the results?)
3. Questions related to oncologists' influence on pathologists' decisions and vice versa	3.1 What impact do you have on the decision to use a particular test? 3.1.1 What influence does the doctor have on the choice of test to be applied? (That is, does the doctor only indicate what to measure or also which test provider to use?) 3.2 When an oncologist sends you a sample for investigation, what information does he or she get back (e.g., test specificity, sensitivity, predictive value)? 3.2.1 Do you feel oncologists are able to work properly with such values?

Note: Every participant was not necessarily asked all of the questions. It depended on the answers given for each question which questions were asked next.

Not every question compiled for the pilot study (Table 4.1) was necessarily asked. What questions were asked depended on answers given to previous questions. Participants were not compensated for their participation.

4.1.3 *Participants*

Nine German pathologists, one woman and eight men, practicing in Berlin, Potsdam, Leipzig as well as in Brandenburg and six US pathologists, three women and three men, practicing in Seattle, Washington, were recruited with the procedure described above. Table 4.2 summarizes the characteristics of the German and US samples³².

With respect to the focus of this thesis, both the US and German samples were considered to be representative of the population of pathologists within their respective health systems. However, significant differences between German and US participants were found regarding age, with participants in the US being generally younger and consequently less experienced. Furthermore it was found that the US sample included more women, and that the samples differed with respect to qualifications, a result of differences in the health education systems of the two countries. (For the US sample, where specialization exists, four of the pathologists were trained as anatomical and clinical pathologists; two were trained as anatomical pathologists³³.)

4.1.4 *Analytical Procedure*

The 15 pathologists' interviews were transcribed and analyzed. The analytical procedure was identical to that used for the oncologists' pilot study (see Chapter III). Each interview protocol was assessed and checked for cues mentioned as being important for the decision to use particular kinds of tests. Again, because the design of the main study limited the number of cues to take into consideration only those cues that were mentioned by at least 50% of each of the respective groups were used in the main study.

³² As was the case for the oncologist pilot study (see Chapter III), I was not able to interview as many US pathologists as German pathologists due to a limited stay in the United States.

³³ Anatomic pathology is related to the science of tissue only, while clinical pathology refers to the sciences of blood or any other body fluids.

Table 4.2: Description of the Fifteen Participating Pathologists

4.2 Sample characteristic	Germany (N = 9)	United States (N = 6)
Age	38 – 62 years (Mean: 51 years; SD: 8.19)	40 – 45 years (Mean: 43 years; SD: 2.10)
Gender	Male: N = 8 Female: N = 1	Male: N = 3 Female: N = 3
Qualification^a	Pathologist: N = 9	Anatomic and clinical pathologist: N = 4 Anatomic pathologist: N = 2
Position^b	Professor/Chief physician: N = 4 Associate professor/ Assistant medical doctor: N = 4	Professor/Chief physician: N = 2 Associate professor/ Assistant medical doctor: N = 4
Years of experience	9 – 32 years (Mean: 21.89 years; SD: 9.09)	10 – 15 years (Mean: 12.50 years; SD: 2.07)
Location/Institute	University/Research clinic: N = 3 Hospital: N = 5 Private practice: N = 1	University/Research clinic: N = 3 Hospital: N = 3 Private practice: N = 0
Analytical procedure used in the laboratory^c	Immunohistochemistry: N = 9 Molecular genetic based: N = 5	Immunohistochemistry: N = 6 Molecular genetic based: N = 3

Note: ^a The education of the US pathologists differed from that of the German participants in terms of the degree of specialization. While the study of pathology in Germany encompasses any possible fields of application, in the United States it is possible to study more focused aspects of it.

^b Pathologists in private practice in Germany in most cases do not hold a specific position in terms of hierarchy, as they usually practice on their own. Therefore the numbers in each position of the German sample do not add up to the total sample.

^c Immunohistochemistry is a basic analytical procedure for tissue used in every lab; molecular genetic-based procedures were used in addition in some of the labs, so the numbers do not add up to the total number of each sample.

4.2.1 Results

“You know, there was the c-kit coming out and it was for testing gastrointestinal stromal tumors’ response to the drug Glivec only. Clinicians got really keen on it, ordered it for all these cancers they did not have a drug to treat at hand in the hope the test would be positive and they could make use of Glivec. Of course, you can tell them that it is not worth doing...that there is no logical reason. But when they want you to do the test, you have to do the test, in this way we are just a service...”

A US pilot study pathologist

All pathologists interviewed had had sufficient experience testing tumors with different methods and with respect to different clinical implications, and all of them reported having sufficient experience with the currently obtainable target tests. Hence, pilot study participants were regarded as being well able to provide sufficient information with regard to the research object in question.

What do pathologists have to decide? The pathologists reported that it is not up to them to decide initially whether a test to answer a diagnostic or therapeutic question is needed—this is the role of the oncologist. The majority of pathologists (N = 10) mentioned that they perform far fewer tests than are available on the market because many of these tests are simply not ordered by the oncologists. However, consistent with the oncologists, pathologists reported that it is usually up to them to decide which testing method should be used to answer a given clinical question posed by an oncologist. For example, in the case of the drug Herceptin, which is thought to be helpful for a specific group of metastatic mamma carcinoma patients (Chapter I), an oncologist requires an answer to the clinical question of whether his/her patient is eligible for the drug. The eligibility for Herceptin can be tested by three methods: one based on immunohistochemistry (IHC), another based on fluorescence in situ hybridization (FISH), and the third based on the technology of chromogenic in situ hybridization (CISH). Usually it would be the responsibility³⁴ of the pathologist to decide which of these three test methods would be the best for his/her laboratory.

How do pathologists decide which test is best? Pilot study pathologists reported they usually obtain information on available tests in scientific meetings, from their respective literature, as well as directly from the pharmaceutical industry. Which test (system) was considered best for their laboratories was commonly determined after numerous trade-offs between the so-called accuracy and reliability of a test on the one hand and the cost effectiveness of the test system on the other. Different considerations were subsumed under cost effectiveness, which participants had difficulty describing. To start with, the *analysis method* the test is based on is the source of most of the further cost-related issues. Very roughly, analyses are divided into

³⁴ In the cases of a hospital or clinic participating in a clinical study carried out by the pharmaceutical industry, degrees of freedom of pathologist's decision is restricted by study protocol, which usually meticulously describes the procedure and test methods to be used in order to ensure comparability between the results obtained from several hospitals and clinics.

microscopic and nonmicroscopic methods. Whereas microscopic methods are standard procedure in laboratories all over the world, and apparently only a microscope—the basic equipment of a lab—is needed, nonmicroscopic methods require specially trained staff as well as expensive apparatus and special equipment, such as ingredients for the test kit, which have to be replenished continuously. Participants reported that nonmicroscopic analyses are comparatively more costly, entailing higher salaries for more highly educated staff, higher outlays for test-specific apparatus, as well as higher operating expenses for purchasing material to run the test kits.

They also looked at analytical methods in terms of what portion of the preparation and performance of the test system requires the pathologist's time versus the laboratory assistant's time. It is worth highlighting that none of the pilot study participants were able to name concrete figures on all of the cost factors. Yet it became obvious in the pilot study that especially for privately run laboratories or hospitals, the cost factor is much more limiting than for those pathology labs that are integrated in university clinics or clinical research centers. Almost all participants coming from hospitals or private laboratories reported that the more expensive molecular genetic methods were not affordable, even while most of them reported a desire to implement such methods in their labs.

Another cue important to most of the interviewees was the so-called *turn-around time* of a test, which refers to the period of time from the moment a tissue sample has been sent to the pathology department to the second the oncologist has the pathology report with the desired test results back. The turn-around time depends on steps of preparation and on the degree of automation of the test system itself. Since a shorter turn-around time is more desired by oncologists, pathologists regard it as being of higher value.

While these mostly cost-related issues represented one side of the “decision equation”, concerns about accuracy and reliability of a test stood on the other. Pathologists reported being aware that no test could deliver results with 100% *accuracy* and *reliability*. However, what rate was acceptable to them would depend on alternatives available and what relative importance the test result had for the medical decision. Asked for accuracy rates of the currently used Her2-new³⁵ test procedures, none of the pathologists knew the correct answer. However, in general, pathologists repeatedly referred to a specificity of some 90%, which they regarded as tolerable and applicable to most of the test methods they used. In terms of

³⁵ Test method(s) for the drug Herceptin.

reliability, the pathologists were not prepared to accept large variations in the results. A value of 90% or more was required for them to opt for a test system. Those pathologists who had one of the HER2-new test methods implemented in their laboratories reported its reliability to be in the range of 93–99% depending on the test system used.

Finally, as was the case in the other studies, pathologists also reported that *expert opinions* influenced their choice to a substantial degree. At the time of the pilot study, at least for the German pathologists, such opinions were most often received from peer literature and scientific meetings, as guideline development was still in a preparation state. Table 4.3 summarizes the decision-impacting cues and their respective number of mentions.

Table 4.3: *Cues Important for the Pathologists' Choice on Applying a Pharmacodiagnostic Test, Presented by Numbers of Mention*

<i>Cues</i>	<i>Germany (N = 9)</i>	<i>United States (N = 6)</i>
<i>Analysis method (cost effectiveness issues)</i>	N = 9	N = 6
<i>Turn-around time</i>	N = 9	N = 6
<i>Cost of the test kit</i>	N = 8	N = 5
<i>Accuracy - Specificity</i>	N = 9	N = 6
<i>Reliability</i>	N = 9	N = 6
<i>Recommendation by experts</i>	N = 8	N = 6

With respect to pathologists' decision-making strategies, they reported making various trade-offs when deciding on a test system for a clinical question that had not been previously addressed in the laboratory. As already mentioned, several cues related to costs as well as to the accuracy of the test (system) are taken into account. However, pathologists were not able to describe what exactly the value of one cue in comparison to the value of another cue is, but merely if one cue value in comparison to another one is regarded being rather of a positive or of a negative value.

No remarkable differences between the German and the US samples were found in terms of the general decision situations the pathologists reported facing in their daily practice or the decisive cues triggering and influencing their decision making. However, US

pathologists appeared to have used a greater variety of tests than their German counterparts. In particular, several of the US pathologists were already aware of the existence of the OncotypDx® test, which was familiar to none of the German interviewees at that time.

4.3 Implications for the Main Study – Do Pathologists Juggle With Cues Weights They Only Feel?

What has been learned about pathologists' decision making, then? Similar to oncologists, pathologists report making trade-offs when facing a decision on implementing a test (system) in their laboratory. It was highlighted that the trigger needed for this decision is a request for a test by oncologists. This means that unlike the oncologists, pathologists are usually confronted with forced choices in their daily practice, and unlike the oncologists, they do not have the option to skip a choice if one or more aspects of the offered options (e.g., cost, reliability, etc.) do not meet their needs, but nonetheless do have to choose one of them. Because of this situation, at first glance trade-offs seem to be more likely to happen despite the assumed task complexity and time pressure clinicians usually face in their daily practice (see Chapter II). However, although all of the participants reported having to balance cost, turn-around time, and accuracy and reliability, it was noticed that participants had limited knowledge of actual costs and accuracy rates of the currently used methods, and instead relied on rough definitions of which value of a cue is of a positive and which one of a negative value to them.

Making use of positive and negative cue information is how Dawes's rule works. Although Dawes's rule is still a compensatory model, it uses unit weights and thus does not involve much computation. In this, Dawes' rule is a fast model and would fit the constraints of pathologists' daily work, while not a frugal one as it still looks up all cues. If pathologists decide in a manner consistent with Dawes's rule, they should always choose the test alternative that has the highest overall score. Consequently the first hypothesis for the pathologists' main study is:

H1: When faced with a decision on two test alternatives, a pathologist will decide in favor of the test alternative with the higher positive score.

In all the interviews it was noticed that the term *specificity* was repeatedly mentioned but not the term *sensitivity*. One university pathologist explained: Imprecise usage of the terms sensitivity and specificity commonly produces confusion in the diagnostic use of sophisticated laboratory test results, such as pharmacodiagnostic tests, since there is not only one sensitivity and one specificity but two of each: Most people familiar with statistics have encountered the concept of diagnostic sensitivity and diagnostic specificity³⁶. In addition, there is *analytical sensitivity* and *analytical specificity*. Analytical sensitivity represents the smallest amount of substance in a sample that can accurately be measured by an assay, while analytical specificity refers to the ability of an assay to measure one particular organism or substance, rather than others, in a sample. High analytical sensitivity does not guarantee acceptable diagnostic sensitivity. That is, a test system's analytical sensitivity and analytical specificity are different in meaning from the system's clinical diagnostic sensitivity and diagnostic specificity. The university pathologist pointed out that most of her nonacademic colleagues, especially those who do not conduct their own research for the development of diagnostic tests, are not very familiar with these differences and in particular not with the terms diagnostic sensitivity and specificity. The specificity the majority of pilot study participants had talked about was the commonly reported analytical specificity. That is, although I was interested in learning more about how the pathologists had dealt with diagnostic sensitivity and specificity within the main study, these terms appeared to be not of major concerns for the majority of participants. Therefore only the analytical specificity was of further interest for the main study.

Furthermore, the university pathologists were found to be in a position to consider molecular genetic-based alternatives to test systems that most of their counterparts employed in hospitals or private practices were not. Due to the research obligations of the university clinic they do have the required know-how for performing the quite sophisticated test procedures and on top of that the required economical capacities for implementing and running the fairly expensive test systems. This puts pathologists coming from the university side in the novel position of having more degrees of freedom when deciding what test system to implement. Some of the pilot study pathologists employed in hospital laboratories stressed their desire to apply new molecular genetic methods but that they simply did not have the

³⁶ *Diagnostic sensitivity* describes the percentage of persons who have a given condition/disorder and who are correctly identified by a test assay as positive for the condition/disorder. *Diagnostic specificity* is the percentage of persons who do not have a given condition who are identified by the assay as negative for the condition.

financial resources and training to do so. This environmental situation should be reflected in the pathologists' choices in the main study. For the university pathologists, a nonmicroscopic test procedure should not carry a negative value, and such tests should be chosen by them more often compared to pathologists not coming from university clinics. However, it is not assumed that the general decision-making strategy is influenced by this facet. Both of these hypotheses have to be tested, of course.

H2a: University pathologists will make more decisions to use nonmicroscopic tests than pathologists coming from hospitals and private practices.

H2b: Pathologists coming from university clinics show no differences in the applied decision-making strategy from their counterparts coming from hospitals and private practices.

The main study's purpose was to test with an acknowledged judgment analysis method whether these hypotheses would stand up to statistical testing. I investigated what decision-making strategy pathologists apply when facing the decision to implement a pharmacodiagnostic test system, what cue or cues are finally decisive for their choice, and whether or not assumed intragroup differences are indeed existent.

4.4 Main study

As it was seen above, pathologists described their decision-making behavior as being consistent with the application of Dawes's rule, that is, they considered themselves to add up positive cue values and subtract negative values. Although Dawes's rule is a fast model, it still looks up all the cues and it is therefore not frugal, which might be an obstacle for clinicians given the demands of their busy and complex daily work (see Chapter II). For this reason I do not assume Dawes's rule will describe and predict pathologists' data better than a fast and frugal model would. Instead, I expect pathologists to make use of fast and frugal decision-making strategies to a proper amount when faced with a decision on pharmacodiagnostic tests. Therefore, it is additionally hypothesized that a fast and frugal model will describe and predict pathologists' test decision-making policies equally as well as a compensatory integration

model. This assumption is based on the existent environmental constraints such as limited time, task complexity, and the general limited computational capacity of humans.

H3a: Pathologists' test decision making will be described equally well by a fast and frugal model as by a complex compensatory integration model.

H3b: Pathologists' test decision making will be predicted equally well by a fast and frugal model as by a compensatory integration model.

4.4.1 *Design*

The same methodological approach as already applied for the oncologists' main study was employed for the pathologists' main study (see Chapter III).

However, different from the oncologists' main study where a single choice was described by one hypothetical test for which participants had to make a 'yes' or 'no' choice, in this study participants had to choose one of two hypothetical test alternatives per choice. Both of the offered test alternatives were designed by almost all cues and the levels identified as being influential for pathologists' choice during the pilot study (see Table 4.3). One of the alternatives within each choice set was held constant, that is, it was the same in all choices. Its cue levels represented an average of reported values of currently used test procedures. The second test alternative of each choice set was this for which the cue levels were manipulated. That is to say, a case within the pathologists' main study means a choice between two test alternatives, out of which one was always held constant and the other was altered for every single case (see Appendix B).

Several cues were found to be decisive for pathologists' decisions on appropriate pharmacodiagnostic tests (systems). These cues were analytical method underlying the test system, turn-around time, ongoing cost of the test kit, specificity of the tests, reliability, and recommendation by experts.

Because pre-testing made it obvious that the pathologists would not be willing to complete more than 10 choices, due to their lack of time, some limitations were set on the choice design as well. To ensure sufficient sample numbers, again a fractional factorial design was used; this retained the cues' orthogonality and ensured the verifiability of the main effects. As in the oncologists' study, two versions of a questionnaire (A and B) were used to

extend the number of cues available to be manipulated. This procedure allowed a design where all of the important cues apart from reliability could be manipulated. Although pathologists reported that reliability is important to them, they also stated that different tests vary only very little on this dimension. For this reason, it was felt that by excluding that cue the least information would be lost. The other five cues (analytic method, specificity of the tests, recommendation by experts, turn-around time, and ongoing cost of the test kit) were all set at two levels, resulting in 16 choice items, 8 per questionnaire version. To ensure a high overlap between the choices presented in the questionnaire and a daily routine situation, cue levels were specified with four pathologists, one from the university side and three from a hospital.

Table 4.4: Cues, their Levels, and Distributions for Versions A and B of the Questionnaire for the Manipulated Test Alternative

<i>Cue</i>	<i>Level</i>	<i>Distribution</i>	
		<i>Part A</i>	<i>Part B</i>
<i>Analytic method</i>	(1) Microscopic	4	4
	(0) Non-microscopic	4	4
<i>Turn-around time</i>	(1) 2 days	4	4
	(2) 1 day	4	4
<i>Specificity</i>	(0) 85%	4	4
	(1) 95%	4	4
<i>Cost of the test kit^a</i>	(1) €/ \$ 280	4	4
	(0) €/ \$ 400	4	4
<i>Recommended by experts</i>	(1) Yes	4	4
	(0) No	4	4

^aEuros and dollars were used for the German and US questionnaires, respectively.

Table 4.5: Cues and Their Levels for the Constant Test Alternative

Cue	Level
<i>Analytic method</i>	(1) Microscopic
<i>Turn-around time</i>	(0) 3 days
<i>Specificity</i>	(0) 85%
<i>Cost of the test kit^a</i>	(2) €/ \$ 200
<i>Recommended by experts</i>	(1) Yes

^aEuros and dollars were used for the German and US questionnaires, respectively.

Each of the cue values of the manipulated test alternative was equally distributed among the set of the 16 choice sets as well as among each version of the questionnaire. The five cues, their levels, and the distribution of their values are shown in Table 4.4. For the constant test alternative, the distribution of its cue value among the choices and the questionnaire versions does not vary. The cue values of the constant test alternative are shown in Table 4.5.

Table 4.6: Questions and Their Possible Responses for Capturing Intragroup Differences and Demographics

Questions	Responses
D1: How old are you?	(1) < 35 (2) 35 - 50 (3) > 50
D2: What is your exact job title?	Open field answer
D3: How many years have you practiced as a professional?	Open field answer
D4: Where do you work?	(1) Academic setting/University clinic (2) Hospital (3) Private laboratory
D5: If you work in a clinic or hospital, which of the titles below describes your position best?	(1) Chief physician/Head of department; Professor (2) Assistant medical director; Assistant professor (3) Assistant doctor; Staff physician
D6: Which methods of analysis do you normally use in your laboratory? (Please check all that apply.)	(1) Immunohistochemistry (IHC) (2) Fluorescence-based methods (e.g. FISH) (3) Molecular genetic methods (e.g. RT-PCR) Micro-array techniques)
D7: Are research projects conducted in your academic or private practice? (Other than clinical trials)	(1) Yes (Please continue with question D8) (2) No (Please continue with question D9)
D8: Are you involved in any of these research projects?	(1) Yes, as principal investigator (2) Yes, as a co-investigator

	(3) No
D9: How many times a year do you attend scientific meetings of your medical society/community?	(1) Less than three times a year (2) Three to five times a year (3) More than five times a year
D10: How many times a year do you attend professional continuing education?	(1) Less than once a year (2) At least once a year (3) At least twice a year (4) More than twice a year

The cover story given to the pathologists told them to imagine the rather common situation that their oncologist wants them to implement a new prognostic test system to better specify cancer patients' prognoses. The result of the tests would be used from then on to help oncologists better determine whether to administer therapy B, which benefits further 5% out of 100 patients, in addition to a standard therapy A, which benefits already solely about 70% of these 100 patients. Then, participants were introduced in detail to all the cues they would have at hand to make that decision. The cover story was regarded as being appropriate for every pathologist in Germany and every anatomically educated pathologist in the United States.

After the cover story was presented participants were told that they would encounter eight pairs of test alternatives and would have to decide for each choice pair which of the two tests they would use to answer the clinical question. Before they made their choices, participants were told that there were no right or wrong answers and they were asked to make their decisions just as they would in daily practice. The presentation of the eight choice pairs of either version A or B of the questionnaire followed. After the choice pairs were presented, five further questions concerning the questionnaire's representativeness were asked (see Chapter III: *Table 3.5*). In addition, because during the pilot study hints were found that differences in age, position, place of employment, application of different analytic methods, involvement in research, and attendance at educational training might have an effect on pathologists' decisions, several questions designed to capture these intragroup differences were asked at the end of the questionnaire. These questions and their possible responses are outlined in *Table 4.6*. The main study was piloted on four participants of the pilot study with only minor adjustments to terminology.

4.4.2 Procedure

The detailed description of the decision situation, the eight choice sets, and the questions concerning the questionnaire's representativeness as well as those for capturing

intragroup differences and demographic issues were presented to the participants either on a webpage or on paper³⁷ (see Appendix B). A short foreword was included that introduced the study as well as the investigator and guaranteed respondents' anonymity. Then, either version A or version B was presented randomly to the participants. The two versions differed only in terms of the choice sets presented and not with respect to any other part of the questionnaire, such as the description of the general situation (cover story). The order of the choices in A was different from the order of the choices in version B to eliminate order effects.

In Germany, two main societies, namely, the Association of German Pathologists and the *German Society of Pathology*³⁸, which are regarded as having access to representatives of the pathologist community, were asked for their support by distributing a letter via email to their members³⁹ providing a short introduction to the investigator and explanation of her research, the request for participation, and the link to the respective webpage (see Appendix B). Furthermore, following personal contact with a pilot study pathologist who was responsible for organizing and conducting professional training for pathologists, 180 paper versions of the questionnaire were distributed. The response rate for the paper version was 18%. For the online version response rate could only be estimated since both societies gave only a vague idea of the number of their members. Furthermore, overlap between the members listed in the directories of the two societies was reported. However, after one society had announced the study to approximately 600 members and before the second society had done so, 46 responses to the online version were counted, a response of about 8%. So the response rate for the German pathologist sample ranged from about 8 to 18%. The sampling frame for the US pathologist sample was obtained from several university clinic and hospital webpages around the country. Over 700 anatomical as well as combined clinical and anatomical pathologists were approached directly via email by the investigator and asked for their participation in the study. The content of that email was nearly identical to that distributed by the German societies to their members. The final response rate was about 16%, which was a satisfying response given that the investigator and her reputation were unknown to the participants.

³⁷ Note that the content of the online version and the paper version of the questionnaire was identical.

³⁸ Berufsverband Deutscher Pathologen and Deutsche Gesellschaft für Pathologie

³⁹ All societies and associations were explicitly asked not to reveal their member lists to the investigator to ensure the highest level of privacy to the participants.

4.4.3 Participants

Ninety-three German pathologists and 108 US pathologists fully completed the questionnaire. While all of the 108 US participants responded to the online version of the questionnaire, data of 32 German participants were gathered by the paper version of the questionnaire. As the initial invitation to participate in the study was distributed to pathologists all over both countries, participants from every region were expected to be captured.

Table 4.7: Demographics and Intragroup Related Values of Main Study Pathologists

Questions	Responses	Germany (N=93)	United States (N=108)
D1: How old are you?	(1) < 35 (2) 35 - 50 (3) > 50	N = 6 N = 50 N = 37	N = 2 N = 64 N = 42
D2: What is your exact job title? ^a	Open field answer	Anatomical and clinical pathologist: N = 93	Anatomical and clinical pathologist: N = 93 Anatomical pathologist: N = 15
D3: How many years have you practiced as a professional?	Open field answer	2 – 42 years (Mean: 20.97; SD: 9.07)	1 – 50 years (Mean: 16.29; SD: 11.16) 2 missing values
D4: Where do you work?	(1) University clinic/ Research center (2) Hospital (3) Private laboratory	N = 20 N = 35 N = 38	N = 86 N = 17 N = 5
D5: If you work in a clinic or hospital, which of the titles below describes your position best? ^b	(1) Chief physician/Head of department; Professor (2) Assistant medical director; Assistant professor (3) Assistant doctor; Staff physician	N = 27 N = 28 N = 4	N = 42 N = 45 N = 17
D6: Which methods of analysis do you normally use in your laboratory? (Please check all that apply.) ^c	(1) Immunohistochemistry (IHC) (2) Fluorescence-based methods (e.g. FISH) (3) Molecular genetic methods (e.g. RT-PCR, micro-array techniques)	N = 93 N = 46 N = 31	N = 108 N = 78 N = 71
D7: Are research projects conducted in your academic or private practice? (Other than clinical trials)	(1) Yes (If YES, please continue with question D8) (2) No (If NO, please continue with question D9)	N = 32 N = 61	N = 82 N = 26
D8: Are you involved in any of these research projects? ^d	(1) Yes, as principal investigator (2) Yes, as a co-investigator (3) No	N = 22 N = 8 N = 2	N = 45 N = 35 N = 2

D9: How many times a year do you attend scientific meetings of your medical society/community?	(1) Less than three times a year	N = 23	N = 56
	(2) Three to five times a year	N = 41	N = 36
	(3) More than five times a year	N = 29	N = 15
			1 missing value
D10: How many times a year do you attend professional continuing education?	(1) Less than once a year	N = 2	N = 2
	(2) At least once a year	N = 6	N = 24
	(3) At least twice a year	N = 9	N = 22
	(4) More than twice a year	N = 76	N = 60

Note: ^a The education of the US pathologists differs from that of the German pathologists in terms of the degree of specialization. While the study of pathology in Germany encompasses any possible fields of application, in the United States it is possible to study more focused aspects of it.

^b Pathologists in private practice in Germany in most of cases do not hold a specific position in terms of hierarchy as they usually do if practicing on their own. Therefore, numbers of the German sample do not add to the total sample.

^c Since immunohistochemistry is a basic analytic procedure for tissue used in every lab and molecular genetic-based procedures could also be used in any of the labs, numbers do not add to the total samples.

^d Numbers do not always add up to the overall sample sizes because not all questions had to be answered.

However, in the US sample mainly email addresses belonging to pathologists working in university clinics were available, which may have caused, this group to be over-presented in the US study. That is, while participants of the German sample are considered to be representative of the community of pathologists in charge of conducting tests on cancer tissue, for the US sample the same cannot be claimed. The samples' demographics and responses to the personal questions designed to reveal intragroup differences are given in Table 4.7.

To start with the German sample (N=93), pathologists' experience in analyzing tumor tissue ranged from 2 to 42 years, with a mean of 20.97 years. 38 participants practiced in private practices, 35 in hospitals and 20 in university clinics. From the ninety and three German pathologists, 27 were chief physicians/professors, 28 assistant medical doctors/associated professors and 4 assistant doctors/staff physician. Every participant reported to apply immunohistochemical procedures in her/his laboratory (N = 93), while only 46 used fluorescence-based method and 31 molecular-based methods. 32 were involved in research, out of whom 22 were principals and 8 coinvestigators.

For the US oncologist sample (N = 108), their work experience ranged from 1 to 50 years, with a mean of 16.29 years. Here, only 17 participants worked in hospitals, while 86 did so in university and research clinics and none in private practices. The majority of the US participants, namely 45, were assistant medical doctors/associated professors, followed by 42 chief physicians/professors and 17 assistant doctors/staff physicians. Again, each of the participants reported to apply immunohistochemical procedures in her/his laboratory (N = 108), while additionally 78 used fluorescence-based method and 71 molecular-based methods.

Of the hundred and eight US pathologists, 82 were involved in research, while 45 were research principals and 35 coinvestigators.

Significant differences between the German and US samples were found with respect to participants' age, title, years of experience, work location, methods of analysis used (namely, the fluorescence-based methods), and the extent of training. Differences in the working location and analysis methods used as well as the extent of training are based on the fact that the US sample had more participants coming from university clinics due to a different acquisition procedure for that sample. The reason for group differences in titles was already explained in the pilot study section.

4.4.4 Analysis Considerations

As for the oncologist study, for this study I was not able to apply a regression model to the data without risking getting distorted results (Tabachnick & Fidell, 1996) as I was faced with the same limitations regarding the choice-to-cue ratio as I was for the oncologists (see Chapter III: *Analysis Consideration*). Therefore, I again decided to apply the two other compensatory integration models with different degrees of complexity, namely, *Franklin's rule* and *Dawes's rule* to individual pathologists' data. It was emphasized that in their daily practice pathologists are faced with a different kind of decision task than oncologists are. The choice task offered to them within the main study reflected this difference in that I presented no binary categorical “yes”/”no” choice to them, but instead two test alternatives from which one had to be chosen. This difference in task, however, required a different noncompensatory competitor, because the Matching Heuristic approach was developed only for such tasks, where one choice option is not described by any cue values, namely the “no”-option. Instead, the *Take The Best* heuristic developed by Gigerenzer and Goldstein (1996)⁴⁰ was chosen as an adequate model.

Table 4.8: Dichotomized Cue Values of Former Polytomous Cues of the Pathologists' Main Study

Cues	Manipulated test alternative A	Constant test alternative B
Analytical method	(1) Microscopic (0) Nonmicroscopic	(1) Microscopic

⁴⁰ The algorithm of Take The Best has parts of its roots in the paper of Gigerenzer, Hoffrage, and Kleinbölting (1991).

<i>Turn-around time</i>	(1) 2 days (2) 1 day	(0) 3 days
<i>Specificity</i>	(0) 85% (1) 95%	(0) 85%
<i>Cost of the test kit</i>	(1) €/ \$ 280 (0) €/ \$ 400	(2) €/ \$ 200
<i>Recommended by experts</i>	(1) Yes (0) No	(1) Yes

Also for the pathologist study, I was not able to determine an outside-criterion with respect to whether the choice made is correct or not as it was in research of e.g., Gigerenzer and colleagues (1996, 1999) (for more details, see Chapter III: *Analysis considerations*). Therefore, in analogy to what is commonly reported as cue validity for models such as Franklin’s rule, I determined for each cue its importance.

For ease of model analysis polytomous cues (*Turn-around time* and *Cost of the test kit*) were dichotomized (see Table 4.8) and for each cue all italicized values were coded as “0” and nonitalicized values as “1”. The dichotomization of cues was based on information received during the interviews and by discussing with participants with whom the pretesting for the main study was done. The finally derived dichotomized cues were used for all of the following models applied on pathologists’ data, Franklin’s rule, Dawes’ rule, and Take The Best. The test decision for each pathologist was modeled on a set of 8 cases.

This model, which was able to make the *best average prediction* (generalization) across all participants, was chosen as the model of the pathologists’ test application policy.

Franklin’s Rule. In this model each cue’s value was weighted according to its influence on the decision. For each case this model multiplied the cue values by their weight of influence (cue’s importance) for each of the two alternatives and summed them. As the alternatives within the cases were made up of binary cues that were coded “0” and “1”, the sum was determined by all cues taking the value of “1” in the cases that discriminate. For the model to make a prediction, it compared the two sums with each other, and predicted a choice for that alternative having the higher sum.

The cue weights were calculated by choosing the value on each cue that had the greatest proportion of a respective test choice in the set of these cases where the cue discriminated between alternatives A and B. The values of a cue are considered to be

discriminative when test alternative A holds a different value of a cue than test alternative B and therefore the difference between A and B is unequal to zero. In a choice where the values of a cue are the same for both alternatives, the difference of these values would equal zero and be nondiscriminative. To provide an example for the cue “analytic method,” whose values were discriminative in 4 of 8 choices: If a participant chose one time test alternative B, which had for that cue the value “microscopic” (1), and three times test A, which had in those cases the value “nonmicroscopic” (0), then the latter value would determine the cue’s importance weight, which was for that case .75. Since the sum was determined by all cues taking the value of “1” in cases where the values discriminate, the value “nonmicroscopic” (0) had to be recoded in “1,” as it was found to be of higher importance for the choice, while the other value (1) had to be recoded in “0”. For the model to make a prediction, it compared the sums of the two test alternatives with each other for each case, and predicted a choice in favor of that alternative having the higher sum.

For the fitting situation, after the model was fitted on all of the eight cases, it was asked to predict the outcome for each choice. If the predicted outcome was the same as the actual choice of a participant, a “hit” was recorded. After all eight choices were examined the hits were summed and set in relation to all the choices (i.e., 8), which led to the final percentage of fit for that participant. Furthermore, the model’s ability to predict unknown cases (generalization) was investigated by using the k -1 cross-validation technique (see also Chapter III). That is, the model was fitted on the first seven cases of a participant in order to predict the unknown eighth case. This procedure was repeated until every case was predicted that way. The final fit and the model’s ability to generalize was defined as described in the previous paragraph. Again, if the predicted outcome was the same as the actual choice of a participant, a hit was recorded.

Dawes’ Rule. In this model cues were unit weighted. For each case this model adds up all of the positive cue values and subtracts all of the negative values for each of the two presented test alternatives. It then compares the two sums with each other, and predicts a choice for the alternative with the higher sum.

Whether a value of the cue was either positive or negative was determined by the proportion of a respective test choice that followed a respective value. To provide an example for the cue “analytic method,” whose values were discriminative in 4 of 8 choices: If a

participant chose three times test alternative A, which had for that cue the value “nonmicroscopic” (0), and one time test B, which had a cue value “microscopic” (1), then the former value would be coded as being of positive value while the latter would be coded as being of negative value for that participant.

As for Franklin’s rule, this procedure was used to determine the model’s fit as well as its ability to predict unknown cases by using the k -1 cross-validation approach (see also Chapter III: *Analysis Consideration*).

Take The Best heuristic. The Take The Best heuristic is based on the assumption that there are environments in which humans rightly or wrongly know which cues are better than others in pointing to a sound choice. It is furthermore assumed that when people are faced with a decision between two or more objects they start by looking up the value of the cue with the highest validity⁴¹. If that cue discriminates, then the search is immediately stopped and the decision is made without taking into account any further cues. If this most important cue does not discriminate, the second most important cue value is looked up, and so forth. Take The Best is noncompensatory in that no amount of contrary evidence of later (maybe unseen) cues can compensate for or counteract the decision made earlier by a previous cue. Additionally, the cue ordering that Take The Best uses is not ‘optimal’ (Martignon & Hoffrage, 2002). People arrive at their cue orders in different ways. One possibility is a genetic default. In many animal species, for instance, we can find a specific order of cues for mate choice. The other possibility is learning; the order of cues can be estimated from the relative frequency with which they had been useful in the past for making a good or correct decision. For example, if we had to decide which of two cities is larger, then the validity of, say, a soccer team cue would be the frequency with which we had found for other pairs of cities that the one with a soccer team was larger.

To determine the fit as well as the prediction of Take The Best to pathologists’ data, the following steps were taken individually for each participant’s data. Each cue’s importance was calculated in the same way as for Franklin’s rule. Cue’s importance was then used to rank

⁴¹ Note different to Gigerenzer’s research I was not able to determine cue validities based on an outside criterion (see Chapter III: *Analysis Consideration*). However, in accordance with what is usually reported as cue validity I determined each cue’s importance instead. The importance was defined by the frequency of how often a respective cue value was followed by a choice for a respective alternative given that the value of the cue discriminated between the two alternatives.

order the five cues, where the first-ranked cue had the highest importance. When cues had equal importance ranks, the rank order was determined by the order in which the information was presented to the pathologist in the respective choice. The rank order determined in which order the model would search through the cases of the respective participants. If, for example, the cue “analytic method” had the highest importance, the model would start by looking up the values of this cue first. If the cue discriminated, the model would immediately predict the alternative that is supported by that cue, without bothering with the other four cues. If the cue did not discriminate, the model would look up the second most valid cue, and so forth. If the prediction was correct a “hit” was recorded. For example: For a decision of pathologist_1, the cue “analytic method” had the highest importance. In the first choice presented to this pathologist, the cue did not discriminate, so the model had to look up the second most valid cue, “recommended by experts.” Here, the cue did discriminate between the two test alternatives offered. Since alternative B had the respective value on that cue, the model correctly predicted the choice of this pathologist—alternative B. The fit of the Take The Best model for each pathologist was the proportion of correctly predicted choices.

Furthermore, the model’s ability to predict unknown choices (generalization) was investigated by using the $k-1$ cross-validation technique; that is, the choice strategy for a respective participant was modeled on the first seven choices and then the model predicted the eighth choice. If the prediction was correct a “hit” was recorded. Then, the model was fitted on the second to the eighth choices and the first choice was predicted. If the prediction was again correct, a “hit” was recorded. This procedure was repeated until all of the eight choices were predicted.

4.4.5 Results

Which one should I take?: Pathologists’ overall choice, and agreement between them. In contrast to the oncologist study, pathologists did not face a yes/no-choice, but were presented with two test alternatives between which they had to choose. Therefore, in the following, I will report on how often they decided in favor of doing the manipulated test alternative (in the following, *test A*) instead of doing the nonmanipulated standard test (*test B*).

Overall, in 26.70% of the cases, pathologists decided to order pharmacodiagnostic test A instead of B, within the German sample 23.90% of the participants did so, and within the US sample 29.10%. Over the set of eight choices, German pathologists opted for test A on

average 1.91 times, while their American colleagues did so 2.33 times. This difference between the two countries had not even a small effect size (Field, 2005)⁴², but was significant ($t(199) = .923, p = .019, r = .06$).

Taking a closer look at each sample, for the German one, the majority who opted for the test A version were pathologists coming from the university side (33.80%; mean over the set of eight choices: 2.70) followed by pathologists working in a hospital (23.90%; mean over the set of eight choices: 1.91), and by pathologists coming from private practices (18.8%; mean over the set of eight choices: 1.50). The picture was quite identical, although not so pronounced for the US sample: here, university pathologists opted in 29.8% of the cases for test A (mean over the set of eight choices: 2.38), with having pathologists from hospitals again in the middle (26.5%; mean over the set of eight choices: 2.12), and, finally, pathologists who worked in private practices (25.00%; mean over the set of eight choices: 2.00). Regarding the proportion of opting for test A, medium sized differences were found within the German sample ($F(2,90) = 7.29, p = .001, \omega = .35$), whereas this difference existed only between the university group and the two others groups, but not between the hospital group and the private practice group themselves. Within the US sample, no difference was found ($F(2,105) = .45, p = .609, \omega = .01$). Investigating differences of the yes-choice behavior between these three groups from each country, a small, but nonsignificant, difference was found for the university groups ($t(104) = .967, p = .336, r = .09$) as well as for the hospital groups ($t(50) = -.64, p = .528, r = .08$), while the difference between the private practice group was medium sized, but nonsignificant ($t(41) = -1.41, p = .302, r = .21$). Figure 4.1 illustrates the respective proportion of yes-choices for test alternative A and B for all pathologists as well as per country.

⁴² Effect size: $r = .10$ (small effect), $r = .30$ (medium effect), $r = .50$ (large effect).

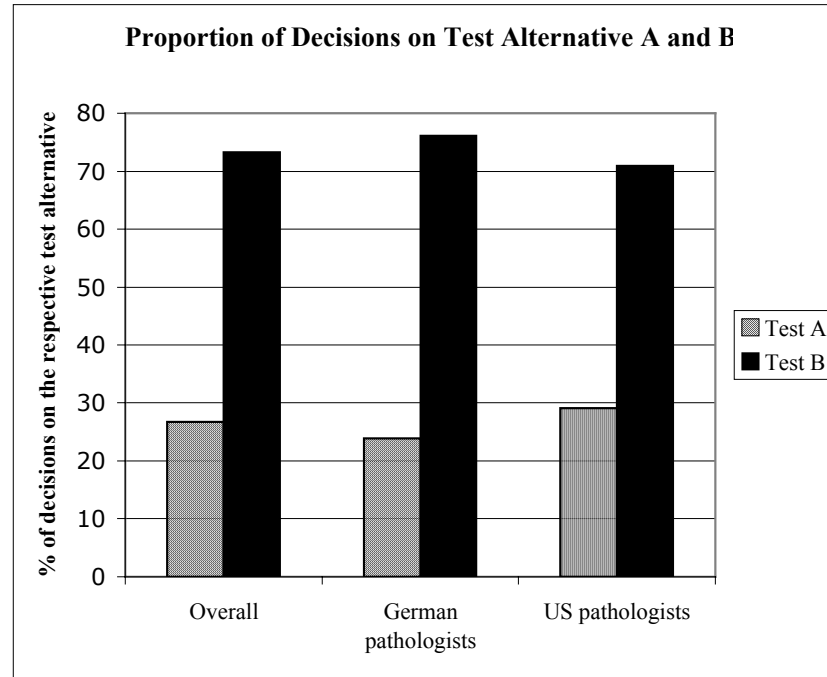


Figure 4.1: *Proportion of yes-choices per test alternative shown for all pathologists (overall) as well as separated for each country.*

The extent of disagreement was again calculated by the percentage of pathologists' disagreement with the modal choice⁴³ on each case (see Chapter III: *Results*). Within each sample, there was disagreement between pathologists regarding the decision to be made on each of the eight cases. The disagreement within the German sample ranged from 0% to 48.94% (mean: 14.09; *SD*: 16.65), and for the US sample fairly identically from 1.82% to 32.72% (mean: 11.98; *SD*: 11.80). This difference in disagreement between the two countries was not significant ($t(30) = .415$, $p = .681$; $r = .08$). German pathologists' disagreement coming from the university side ranged from 0% to 41.67% (mean: 11.44; *SD*: 15.22), while for their US counterparts this ranged from 2.22% to 33.33% (mean: 12.64; *SD*: 11.17). For pathologists coming from hospitals, disagreement between those coming from Germany ranged from 0% to 47.62% (mean: 12.20; *SD*: 15.26) and within the US sample from 0% to 50.00% (mean: 9.72; *SD*: 17.00). For pathologists from private practices, in Germany their disagreement ranged from 0% to 38.89% (mean: 10.31; *SD*: 15.16), while there was no

⁴³ The modal choice was defined by the choice of the majority of the group on this respective case.

disagreement within the US sample. The disagreement between the three German groups did not differ significantly from each other ($F(2/45) = .06, p = .939, \omega = .02$).

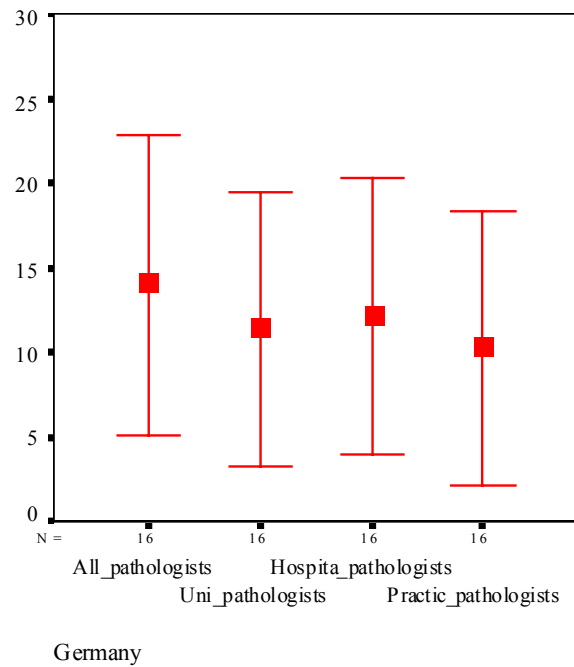


Figure 4.2: Disagreement (mean: 95% CI) over the 16 cases (8 per version) of the German pathologist sample for all pathologists (*All_pathologists*) as well as separated for pathologists coming from university clinics (*Uni_pathologists*), hospitals (*Hospita_pathologists*), and private practices (*Practic_pathologists*).

As for the US sample, no disagreement between the private practice groups was found, and, therefore, the variance of this group was zero, the difference in disagreement was only investigated between the university and the hospital group. For these two groups, differences in disagreement proved to be small and nonsignificant only ($t(30) = .57, p = .570, r = .10$). With respect to the disagreement of these three groups across countries, neither disagreement within the group of university pathologists ($t(30) = -.26, p = .801, r = .05$) nor those within pathologists coming from hospitals ($t(30) = .43, p = .667, r = .08$) showed to be significantly different from each other. As there was no disagreement found for the US private practice pathologists, no search for meaningful differences between the US and the German sample was made. Figures 4.2 and 4.3 display the mean of disagreement, and the 95% confidence interval over all pathologists as well as separated by work location per country.

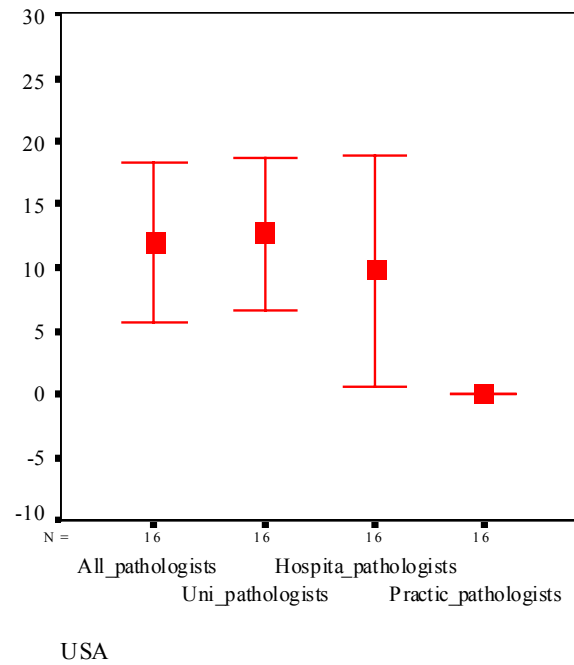


Figure 4.3: Disagreement (mean: 95% CI) over the 16 cases (8 per version) of the US pathologist sample for all pathologists (*All_pathologists*) as well as separated for pathologists coming from university clinics (*Uni_pathologists*), hospitals (*Hospita_pathologists*), and private practices (*Practic_pathologists*).

Happy fitting?: Description of pathologists' test decision policy. In the main study section, it was hypothesized that a fast and frugal model will provide a comparably good fit to the pathologists' test decision-making data as does a complex compensatory integration model (H3a). Due to the limitations of the design, I opted for Franklin's rule and Dawes' rule as representatives for compensatory integration models and on the Take The Best Heuristic as a representative of a fast and frugal model.

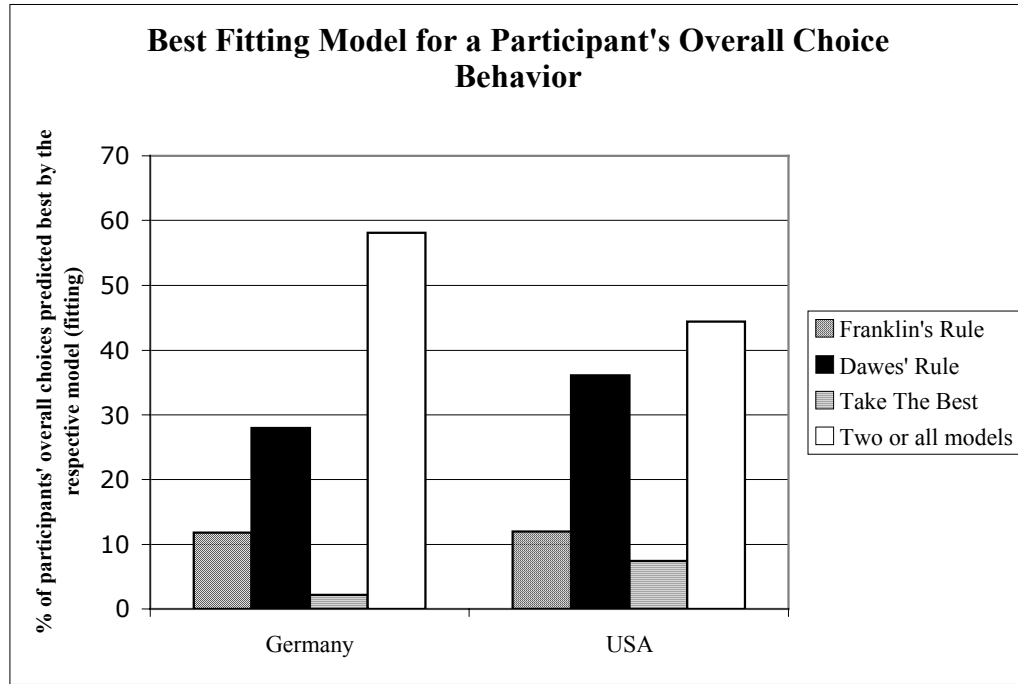


Figure 4.4: Proportion of the best fitting model for a participant's overall choice behavior provided by either Franklin's rule, Dawes' rule, Take The Best, or by two or all models for the German and the US pathologists.

With respect to which model, out of the three applied, fitted a single participant's choices best, that is, counted most hits over all choices of a participant, Franklin's rule provided the best fit for 11.80% of the German and 12.00% of the US participants, Take The Best fitted 2.20% of the German and 7.40% of the American participants best, while Dawes' rule proved to fit 28.00% of the German and 36.10% of the US participants best. These findings do not necessarily have an implication on the general achievement of a model to fit data well, on average. A model can have an achievement of 100%, that is, fit all participant's choices correctly, but share this overall fit with another model. Thus, the previous results give information about how sensitive a model is to fit a single participant's choices on average better than the other two models, that is, to share least best overall fits with other models. For the remaining 58.10% of the German and 44.40% of the US participants, either two or all three models provided an equal overall fit for a participant. Figure 4.4 outlines these findings. Investigating for any intragroup differences, with respect to which model fitted data best, calculating a Chi-square test was again not possible within the two samples, as expected cell sizes for either Franklin's rule or Take The Best, or for both, were lower than 5.0, which violated a major assumption of the Chi-square test procedure. When putting those participants

together who were best fitted by Dawes' rule and comparing these participants to the participants belonging to the "Non-Dawes-rule" group, no significant differences were found.

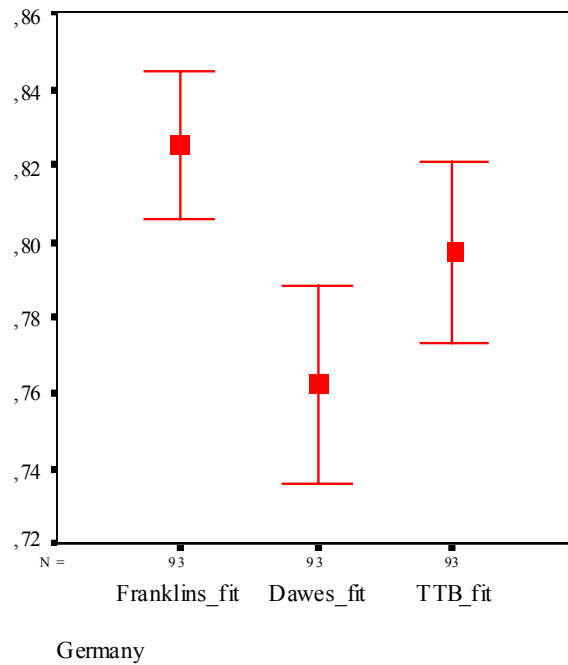


Figure 4.5: Results of the average fit (mean: 95% CI) for Franklin's rule, Dawes' rule, and Take The Best for the German pathologists.

Across all participants, the average fit of Franklin's rule was 82.53% for the German sample (range = 62.5%–100%) and 79.98% for the US sample (range = 62.5%–100%). For Dawes' rule, the average fit of the model was 76.21% (range = 50%–100%) for German pathologists and 74.77% (range = 55%–100%) for their US counterparts. Finally, for Take The Best, here the average fit for the German sample was 79.70% (range = 60%–100%) and for the American sample 76.39% (range: 62.5%–100%). Figures 4.5 and 4.6, respectively, exhibit the mean of the average fit as well as the 95% confidence interval per country. Within the German sample, differences in degree of average fit for the models indicated significant differences between each of the models, while Franklin's rule and Dawes' rule differed to a small extent and Dawes' rule and Take The Best to a marginal extent from each other ($F(2, 184) = 9.08, p = .000, \omega^2 = .12$). For the US sample, the same picture was observed ($F(2, 214) = 8.52, p = .000, \omega^2 = .14$). Between both countries, differences in average fit were nonsignificant and small for Franklin's rule ($t(199) = 1.92, p = .056; r = .13$) as well as for Dawes' rule ($t(199) = 1.10, p = .273; r = .08$), while it was significantly different, although small, for the Take The

Best groups ($t(184.12) = 2.13, p = .034; r = .16$) with a slightly higher mean fit by this model for the German sample.

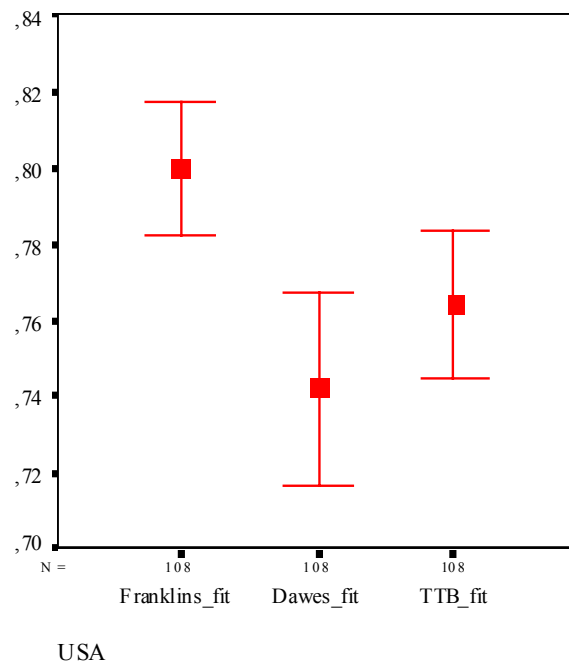


Figure 4.6: Results of the average fit (mean: 95% CI) for Franklin's rule, Dawes' rule, and Take The Best for the US pathologists.

With Hypothesis 3a, it was stated that a fast and frugal model, such as Take The Best, would fit the pathologists' data as equally well as compensatory models, such as Franklin's rule and Dawes' rule. For both countries, Take The Best performed fairly well, given its parsimony in cue use, and even outperformed one of the compensatory competitors, namely Dawes' rule. However, it was not able to perform as well as the compensatory Franklin's rule. Given these findings, Hypothesis 3a can only be confirmed when comparing Take The Best with Dawes' rule, but not with Franklin's rule.

How general are our models?: Generalization performance of the three models. With respect to the generalization performance of the three models, it was hypothesized, additionally, that a fast and frugal model, such as Take The Best, would also provide an equally good prediction to unknown data (generalization) than one of the more complex compensatory models, such as Franklin's rule and Dawes' rule (H3b).

When looking at models' generalization performance, Dawes' rule provided the best prediction to unknown cases, that is, counted most hits for a single participant, compared to the other two models, for 34.40% of the German and 40.70% of the US participants. In this respect, Franklin's rule performed best for none of the German participants, while it did for 2.80% of the American participants. Take The Best also showed to be unable to outperform one of the other two models—it predicted neither for the German nor for the USA sample a participant's unknown cases better than the others.⁴⁴ For the remaining 65.60% of the German and for 56.50% of the US participants either two or all three models provided an equal overall prediction. Figure 4.7 shows the respective results.

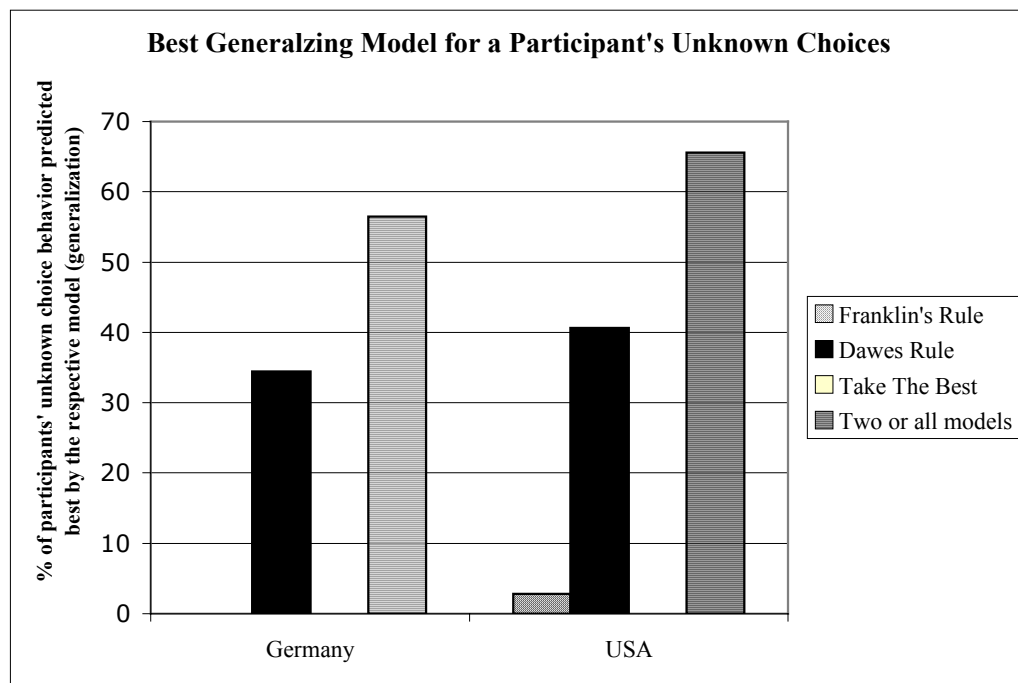


Figure 4.7: Proportion of the best generalizing model for a participant's overall choice behavior provided by either Franklin's rule, Dawes' rule, Take The Best, or by two or all models for the German and the US pathologists.

Since a single Franklin's rule group as well as a Take The Best group from each country was either not existent or only very small, I compared, with a Chi-Square test, every participant not best predicted by Dawes' rule with those who were, with respect to any

⁴⁴ Note, that low proportions do not necessarily imply that the respective model has a low prediction performance, on average. Instead, results refer to a model's ability to predict the overall choice for a single participant better than the other two models, and in that it does not share an equal performance with another model.

intragroup differences. For none of the questioned issues was a meaningful difference between these two groups (Not-Dawes vs. Dawes) found.

Across all participants, Franklin's rule provided the best average prediction of unknown cases for 76.88% (range = 50%–100%) of the German participants and 73.72% (range = 50%–100%) of the US participants. The Dawes' rule average, with respect to this, was 76.21% (range = 50%–100%) for the German sample and 74.19% (range = 50%–100%) for the US sample, while Take The Best provided an average prediction of 74.87% (range = 37.5%–100%) for the German sample and of 71.53% (range = 50%–87.5%) for the US sample. Figures 4.8 and 4.9, respectively, exhibit the mean of the average prediction per model as well as the 95% confidence interval of each per country.

Within the German sample, differences in the degree of average prediction for each model were not significant ($F(2, 184) = .776, p = .462, \omega^2 = .05$). For the US sample, I found a quite identical picture ($F(2, 214) = 2.05, p = .131, \omega^2 = .14$).

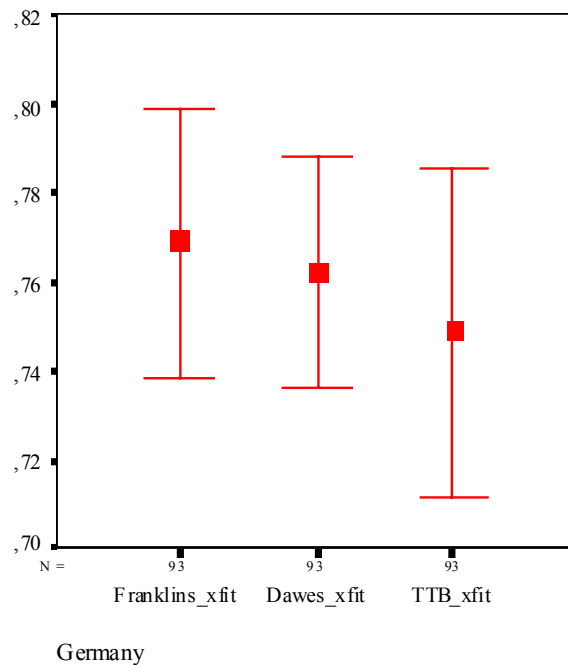


Figure 4.8: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Take The Best for the German pathologists ($N = 93$).

Comparing each respective model between both countries, for each model comparison, I found that their average generalization ability differed only slightly and not significantly from each other (Franklin's rule: $t(199) = 1.68, p = .094, r = .12$; Dawes' rule: $t(199) = 1.10,$

$p = .273$; $r = .08$; Take The Best: $t(199) = 1.44$, $p = .151$; $r = .10$). Given these findings, I can confirm for both countries Hypothesis 3b, that the fast and frugal model, such as Take The Best, is equally able to predict the unknown data (generalization) of pathologists' test decision making more than complex compensatory models, such as Franklin's and Dawes' rule.

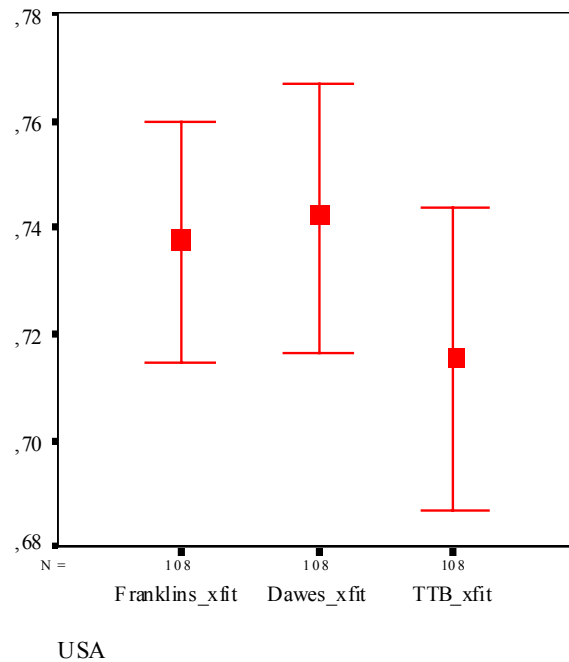


Figure 4.9: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Take The Best for the US pathologists (N= 108).

Taking all or one?: Pathologists' cue use. Within the generalization setting, I found that all three models predicted pathologists' test decisions equally well. Since due to this fact it cannot be clearly stated which model out of the three best reflects the pathologists' decision-making policy of each country, for each of them the respective cue(s) used will be outlined.

For *Franklin's rule*, it was outlined that this model assumes that people take all cues into account and put corresponding weights of importance on each of them. With respect to the generalization setting, for the German sample, the cue with the highest importance was *recommendation by experts* (.98), followed by *analytic method* (.81), *cost of the test kit* (.76), the *turn-around time* (.76), and finally the test's *analytic specificity* (.66). For the US sample, the same cue importance order was found with only slightly different importance values, compared to the German sample: *recommendation by experts* (.96), followed by *analytic*

method (.75), cost of the test kit (.71), the turn-around time (.70), and the test's analytic specificity (.58). Results for both countries are shown in Figure 4.10.

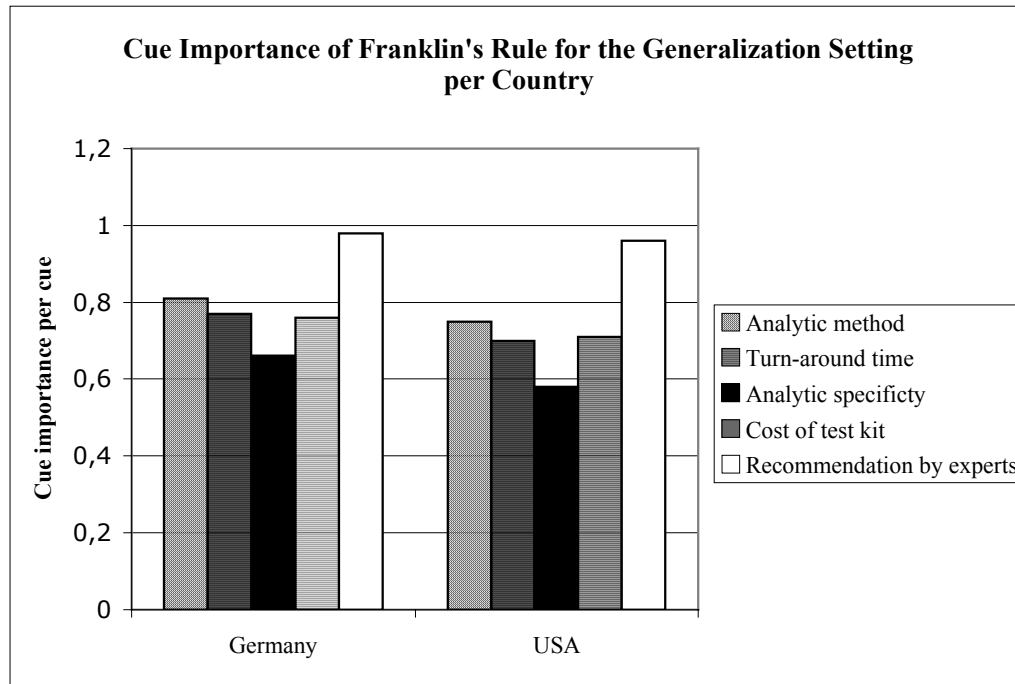


Figure 4.10: Cue importance of Franklin's rule for the generalization setting per country.

For *Dawes' rule*, I mentioned (see Analysis consideration) that this is a unit weight model, which simply works by adding up the number of positive cue values and subtracting the number of negative cue values. On the basis of model's analysis each cue's unit weights were found to be the same as set upfront, therefore, the respective positive (1) and negative (0—italic) cue weights for each value are the same as shown in Table 4.8. While test alternative B consistently had the same three positive and two negative cue values, and, therefore, always had an overall score of +1, test alternative A can range from having four positive cue values and one negative (overall score +3) to only one positive cue value and four negative ones (overall score—3).

Take The Best assumes that when people are faced with a decision between two or more objects, they first start to look up the cue value with the highest validity⁴⁵ (best cue) of at least two objects. Cue importance weights, which determined the rank order, were identical to those reported for Franklin's rule (see Table 4.10). If it discriminates between the two objects,

⁴⁵ In the case of this study this was determined as cue's importance instead.

that is, one object has the respective value and the other does not, then the search is immediately stopped and the decision is made without taking any further cue into account. If this most important cue does not discriminate when both alternatives have the same cue value, the second most important cue value is looked up, and so forth. For both samples, the number of cues used over all eight cases was calculated for each pathologist. Across all German pathologists, the mean number of cues used was 1.48 (range: 1–3; *SD*: .23), while for the US pathologists it was 1.52 (range: 1–3; *SD*: .18). For the German choices, Take The Best predicted choices in 63% cases by using one cue, 25.5% by using 2 cues, and in 11.4% cases by using three cues. Comparable findings were observed for the US sample. Here, Take The Best predicted choices in 60.6% cases by searching one cue only, in 26.9% cases by searching two cues, and in 12.5% by searching three cues. However, participants differed with respect to what cues they applied. For the German sample, the most often used cues were *recommendation by experts* (95.9%), *analytic method* (92.5%), and *costs of the test kit* (88.2%), which was followed by rather rarely applied cues, such as *turn-around time* (7.5%) and *analytic specificity* of the test (2.2%). For the US sample, identically, the most frequently used cues were also here *recommendation by experts* (97.2%), followed by *analytic method* (92.6%), and *costs of the test kit* (84.3%), while again *turn-around time* (11.1%) and *analytic specificity* of the test (3.7%) were found to be rarely used.

Reported representativeness of the questionnaire. With respect to the felt representativeness of the questionnaire, the majority of both the German as well as the American pathologist sample regarded answering the questionnaire easy and the described situation as straightforward. In addition, they found that the offered decision criteria in the scenarios also were the criteria they would use when faced with such a decision. Only a minority of pathologists from each sample requested further information (see next paragraph—*Requests for further information*). The time participants needed to work through the scenarios ranged from 2 to 20 minutes (mean: 7.93; *SD*: 3.65) for the German sample and from 1 to 15 minutes (mean: 6.92; *SD*: 3.20) for the US participants. Investigating for differences between the two samples with a Chi-square test for the variables *F1* to *F3*, showed only small and nonsignificant differences for each of them. However, samples differed significantly, although small, with respect to the time needed to work through the scenarios ($t(197) = 2.10, p = .037; r = .15$). Table 4.9 exhibits all respected values in detail.

Requests for further information. A total of 12 requests for more information within the German and a total of 19 within the US sample were made by pathologists when asked if they missed information within the scenarios that they would normally use for decisions (F4). For the German sample, 33% of these 12 pathologists requested information about how fast the oncologists would need the results returned. Another 33% of the German requests were for more detailed information about cost issues, for example, if experienced staff might be needed for carrying out the test. Seventeen percent asked for information about sensitivity, and a further 17% wanted to know more about the type of cancer they were to diagnose with either test.

Table 4.9: Values of Reported Representativeness of the Questionnaire

Question	Germany (N = 93)	USA (N = 108)
F1	(1) N = 72–77.4% (2) N = 20–21.5% (3) N = 1–1.1%	(1) N = 86–79.6% (2) N = 20–18.5% (3) N = 2–1.9%
F2	(1) N = 78–83.9% (2) N = 14–15.1% (3) N = 1–1.1%	(1) N = 81–75.0% (2) N = 27–25.0% (3) N = 0
F3	(1) N = 81–87.1% (2) N = 12–12.9% (3) N = 0	(1) N = 89–82.4% (2) N = 19–17.6% (3) N = 0
F4	See below (requested information)	See below (requested information)
F5	1–30 minutes (mean: 8.94; SD: 4.12)	2–20 minutes (mean: 8.61; SD: 3.70)

For the US pathologists, of those 19 requesting further information, 32% were interested in sensitivity information. Twenty-one percent of these pathologists wanted to know more about the type of cancer they were to diagnose with either test. Another 21% requested details on how fast the ordering oncologist would need the results returned, while 26% desired more information on cost issues, for example, the compatibility of either test with the already implemented tests in the laboratory.

Findings beyond fitting and predicting. Within the pilot study, section I moreover proposed some further hypotheses beyond fitting and predicting. In the following, the results for these hypotheses will be presented.

Within the pilot study, pathologists described their decision making in the same way that Dawes' rule proceeds. If this were true, it was hypothesized that they would more often

choose this test alternative out of the two presented, which had the higher positive score. This idea was formulated in Hypothesis 1. In the generalization setting (see Hypothesis 3b) Dawes' rule has shown that it was as equally able to predict unknown pathologists' data as the other two models. For these reasons, I wanted to find out if the assumption formulated with Hypothesis 1 would also hold. In order to test this hypothesis, all of the cases of test A were taken, which had a higher positive score⁴⁶ than the nonmanipulated test alternative B. For this analysis, I also included cases of test A, which had an equal positive score with test B, when the cue *analytic method* had the value *nonmicroscopic* within test A. I once coded that cue value as zero (see Table 4.8), since those pathologists coming from hospitals and private practices reported it to be more of a negative utility for them. However, since this cue value was found to be of positive value for the pathologists coming from the university side, I decided on including these cases. For the German sample, for the cases of test A having a higher positive score than test B, opting for these cases ranged from 51.06% to 78.72% (*mean*: 65.06, *SD*: 15.52), while for the US sample it even ranged from 67.27% to 94.34% (*mean*: 79.24, *SD*: 12.21). For the cases of test A, which had an equal or lower positive score than test B, for the German sample, it was observed that here 0% to 68.09% (*mean*: 10.16, *SD*: 19.33) of the participants opted for test A, while for the US sample this was 1.82% to 69.09% (*mean*: 12.23, *SD*: 19.33) of the participants. Figure 4.11 illustrates the range and mean in opting for test A, given a higher or equal/lower positive value, respectively, per country.

When comparing the cases with a higher positive score with those having an equal or lower score within the German sample, a large sized significant difference in the proportion for deciding to carry out test A in favor of these cases was found, where the positive score was higher than that of test B ($t(92) = 7.12; p = .000, r = .59$). The same was observed for the US sample ($t(107) = 8.00; p = .000, r = .61$). Across both countries, I found a small sized and significant difference with respect to the proportion of opting for test A when it had a higher score than test B, with American pathologists opting more frequently for test A than their German counterparts ($t(181.77) = -2.76; p = .006, r = .20$). However, only a small and nonsignificant difference in choice behavior for the cases of test A, which had an equal or lower positive score, compared to test B between the two countries, ($t(199) = -1.03; p = .304, r = .07$) were observed.

⁴⁶ These cases were: 3A as well as 8B. Cases with the “nonmicroscopic” cue value were: 4A and 7B.

Given these results, I confirm Hypothesis 1 for both countries that pathologists decide significantly more on test A cases having a higher positive score than test B, compared to cases of test A having an equal or lower positive score than test B. Apart from the fact that these findings confirm the hypothesis, they furthermore verify that the positive and negative cue weights, which were determined upfront by pilot study findings, prove to be true for the general population of pathologists.

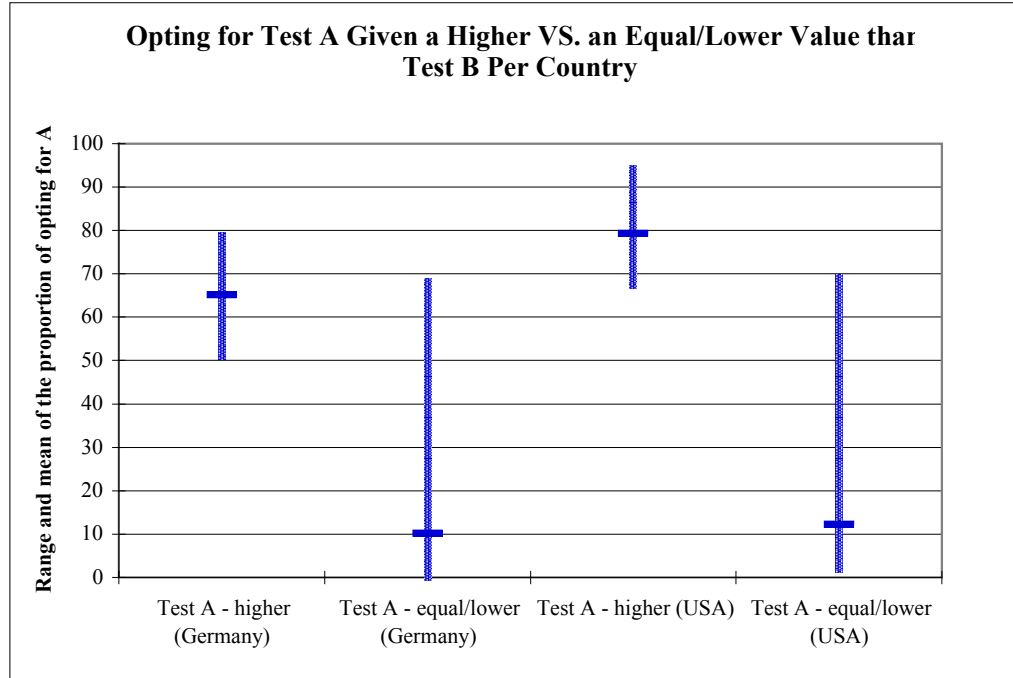


Figure 4.11: Range and mean for the proportion in opting for test A, given a higher positive value and an equal/lower positive value, respectively, than test B, per country.

I furthermore hypothesized that choices of pathologists coming from the university side are characterized by more decisions on nonmicroscopic tests than those by pathologists coming from hospitals and private practices (Hypothesis 2a). In order to test this hypothesis, I first recoded all participants coming from hospitals (2) and private practices (3) into the same value (2), while university pathologists kept the value (1) (see Table 4.6: *Question D4*). Then, I compared the proportion of test A choices between these two pathologist groups for all the case pairs, where test A had the respective cue value.⁴⁷ For the German sample, the proportion of opting for the nonmicroscopic test A, in cases of university pathologists ranged from 0% to 100% (*mean*: 34.90; *SD*: 36.46), while for those who were not from the university this ranged

⁴⁷ These cases were: 2A, 4A, 6A, 8A, 1B, 3B, 5B, and 7B.

from 0% to 50% (*mean*: 14.77; *SD*: 19.56). Within the US sample, university participants' choices for test A ranged from 2.22% to 75.61% (*mean*: 25.89; *SD*: 29.95), and from 0% to 60% (*mean*: 19.79; *SD*: 27.95) for participants from hospitals and private practices. Figure 4.12 exhibits the respective proportion values per country. Within the German sample, university and nonuniversity participants' proportion to opt for a nonmicroscopic test was significantly and largely sized different from each other, and higher for those participants coming from university clinics ($t(24.86) = 3.42, p = .001, r = .56$). However, for the two US pathologist groups, only a small and nonsignificant difference in opting for a nonmicroscopic test was observed ($t(106) = 1.45; p = .075, r = .13$). Comparing the proportion of choices for a nonmicroscopic test A across both countries, neither between the German and the US university groups ($t(104) = 1.58; p = .188, r = .15$) nor between the nonuniversity groups ($t(93) = -1.10; p = .273, r = .11$) was a meaningful difference observed. In the light of these findings, which proved a significantly higher proportion of choice for a nonmicroscopic test from university pathologists than from nonuniversity pathologists within the German sample only, I consider the Hypothesis 2a to be confirmed for Germany, but not for the USA.

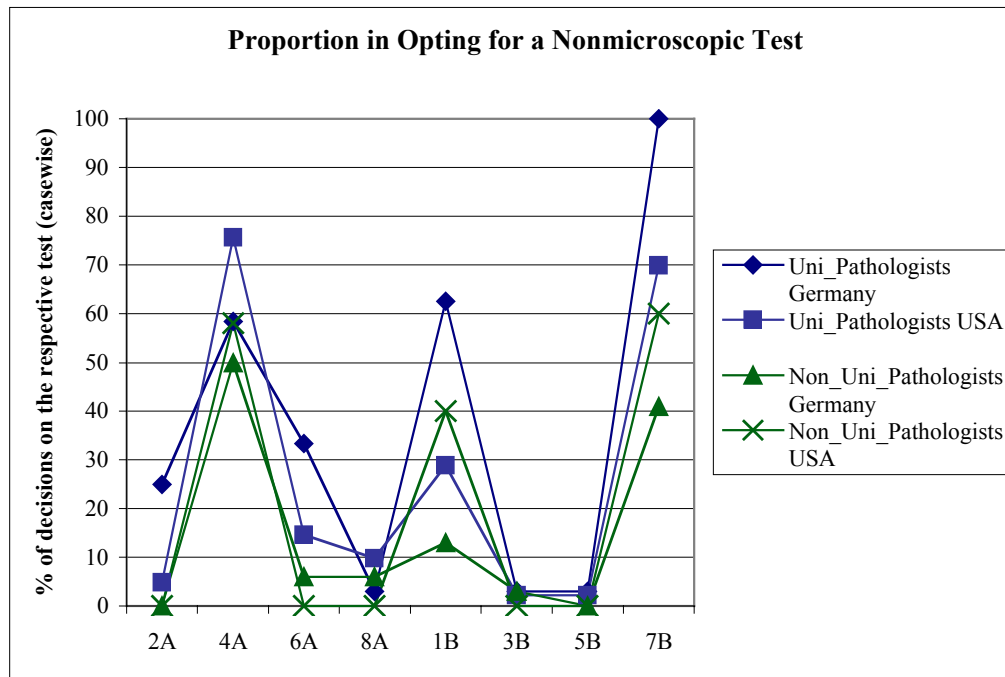


Figure 4.12: Proportion in opting for a nonmicroscopic test per country depending on the location of work of the pathologists (H2a).

With the last Hypothesis 2b, it was proposed that pathologists coming from university clinics show no differences regarding the applied decision-making structure, compared to their counterparts coming from hospitals and private practices. For this study, all of the models were found to provide a quite equal average prediction for both of the pathologist samples. In order to investigate Hypothesis 2b, all university clinic participants (1) and all those who were not from the university clinic were chosen (2), and were compared with respect to the average prediction each of the three models provided. The mean of the average prediction and the standard deviation for each model can be seen in Table 4.10 for the German sample, and in Table 4.11 for the US sample.

Table 4.10: Mean and Standard Deviation for Each Model Exhibit for German Pathologists Coming From University Clinics (1) As Well As for Pathologists Coming From Hospitals and Private Practices (2)

<i>Model</i>	<i>Location^a</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>
<i>Franklin's rule</i>	<i>1</i>	20	68.13	19.23
	<i>2</i>	73	79.28	12.36
<i>Dawes' rule</i>	<i>1</i>	20	73.75	11.40
	<i>2</i>	73	76.88	12.96
<i>Take The Best</i>	<i>1</i>	20	62.50	23.99
	<i>2</i>	73	78.25	14.29

^a Location 1 = Pathologists coming from university clinics; Location 2 = Pathologists coming from hospitals and private practices.

Table 4.11: Mean and Standard Deviation for Each Model Exhibit for US Pathologists Coming From University Clinics (1) As Well As for Pathologists Coming From Hospitals and Private Practices (2)

<i>Model</i>	<i>Location^a</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>
<i>Franklin's rule</i>	<i>1</i>	86	73.26	12.30
	<i>2</i>	22	75.57	9.82
<i>Dawes' rule</i>	<i>1</i>	86	73.11	13.76
	<i>2</i>	22	78.41	10.34
<i>Take The Best</i>	<i>1</i>	86	70.93	15.39
	<i>2</i>	22	73.86	12.74

^a Location 1 = Pathologists coming from university clinics; Location 2 = Pathologists coming from hospitals and private practices.

For the German sample, while between the two respective groups no difference was found for Dawes' rule ($t(91) = -.98, p = .329; r = .10$), for Franklin's rule as well as for Take The Best it was found that both were less able to predict pathologists' data coming from the

university side than those of pathologists not coming from there ($t(91) = -3.14, p = .002; r = .31$ and $t(91) = -2.80, p = .010; r = .28$, respectively). While for Franklin's rule this difference was of medium size, for Take The Best it was of a small size. For the US sample, no difference in the average prediction of these three models between the two groups was found (Franklin's rule: $t(106) = -.82, p = .418; r = .08$; Dawes' rule: $t(106) = -1.69, p = .095; r = .16$; Take The Best: $t(106) = -.82, p = .412; r = .08$). Thus, the Hypothesis 2b, which stated that pathologists coming from university clinics show no differences regarding the applied decision-making structure, compared to their counterparts coming from hospitals and private practices, cannot be confirmed for the German pathologist sample. But this can be confirmed for the US pathologist sample.

4.5 Summary & Discussion

The aim of the pathologist main study was to learn about their decision-making strategies and cue use when faced with a decision on implementing a cancer treatment related pharmacodiagnostic test within their laboratory. Data were derived from eight presented hypothetical cases, which were developed on the basis of findings from a pilot study. To answer the questions of this main study I confronted the data with two compensatory, integrative models, namely Franklin's rule and Dawes' rule, and one noncompensatory model, Take The Best. I investigated the relative ability of these three models to describe and predict decisions made by German and US pathologists in response to systematically varied cases each offering two pharmacodiagnostic test alternatives between to choose.

For both German as well as US pathologists it was found that the majority of decisions were made in favor of the nonmanipulated test alternative B, which was supposed to present an equivalent of currently used standard procedures. While German pathologists decided in 23.9% of the cases on having manipulated test alternative A instead of B, US pathologists did so in 29.1% of the cases. Nevertheless also pathologists disagreed to some extent as to the decision to be made on the same case, while they did remarkably less ($M: 13.04; SD: 14.24$ – for both) than oncologists did ($M: 23.75; SD: 17.55$ for both – see Chapter III). The extent to which they disagreed did not differ between location and country, except for US pathologists coming from private practices. Here no disagreement was found, which was expected to be due to the few members belonging to that group within this study ($N = 5$). Additionally, German as well as US pathologists' test decisions were best described by the compensatory,

integrative Franklin's rule. However, for the generalization setting the compensatory integrative models Franklin's rule and Dawes' rule as well as the noncompensatory model Take The Best made equally good predictions to unknown cases. Furthermore, when investigating pathologists' decisions under the assumption of Dawes' rule it was found for both German and US participants that they opted significantly more often for such test alternatives, which had the higher positive score than the respective counterpart within the choice set. Finally, German pathologists coming from university clinics opted more often on having nonmicroscopic tests than their nonuniversity-side counterparts did, a difference that was not noticed for the US sample. Moreover, for German pathologists coming from the university side it turned out that neither Franklin's rule nor Take The Best delivered an average prediction to their data as good as for the German pathologists not coming from university clinics, while Dawes' rule did. For US pathologists it was not found that those who were coming from the university side were different from those who were not with respect to the applied decision-making strategy.

Taking a look back at the choices made in favor of test alternative A or B it was noticed that only around one fourth and one third, respectively, of pathologists had decided on having test alternative A instead of B. One might wonder what the reason was for that. Had alternative A been too unattractive compared to test alternative B or vice versa, test alternative B too attractive compared to A? I am quite confident that I considered all things important to make it a fair competition between the two alternatives. Due to the forced choice⁴⁸, which I presented to participants I had to ensure that at least one test alternative would represent current test procedures. Test alternative B was supposed to present this current standard within each choice, while test alternative A was supposed to cover the range of upcoming pharmacodiagnostic tests. The only compromise, which was made in this respect, was to choose for the standard version a higher cost level than the average would have been and for test A rather lower ones as price recommendations of upcoming pharmacodiagnostic tests would suggest⁴⁹. This happened with the aim of not making test alternative A an alternative completely out of scope compared to B by cost figures pathologists are not yet used to. To

⁴⁸ No 'not doing a test at all' option was included, which was found within the pilot study to mirror pathologists' reality given an oncologist wants an answer to a clinical question with a test.

⁴⁹ For instance, the pharmacodiagnostic test *OncotypDx*[®] test mentioned in Chapter I was launched with a price recommendation of approximately \$3,000. This was reported to be a significant reason for not having implemented the test system yet.

proceed that way is highly recommended by the respective literature (e.g. Scott, 2002; Louviere, Hensher, & Swait, 2000) working with identical methods such as those used here. But I cannot make things significantly more attractive than they are in reality. And without doubt, it is the characteristic of many realistic decisions that favorable and nonfavorable features co-vary over the offered alternatives (Abelson et al, 1985), for example better quality is accompanied by higher cost. So it was in the presented choice sets and therefore I dare to draw the conclusion based on the findings that by now upcoming pharmacodiagnostic tests in certain cases are not that favorable to pathologists than current standard procedures.

Additionally, I observed that all of the three models predicted pathologists' decisions equally well, and Dawes rule's performance was further fostered by hypothesis one, whose findings showed that both German and US participants opted significantly more often for the test alternative that had a higher positive score. Without any doubt, such overlaps as observed here do have implications for policy capturing. Suppose a researcher would apply Franklin's rule exclusively (or in case of a better case-to-cue ratio the comparable regression equation), s/he would never even detect that the same prediction was derived by a faster and more frugal model and therefore conclude that a compensatory, complex model such as Franklin's rule or regression, respectively was a perfect tool in describing and even predicting the respective choice behavior. In case s/he had applied a different model as I did, s/he would still have to solve the problem of how to deal with them having quite the same predictions. One way to solve that problem is to make use of the principle called Occam's razor (also spelled Ockham's razor). The principle states that the explanation of any phenomenon should make as few assumptions as possible, eliminating, or "shaving off", those that make no difference in the observable predictions of the explanatory hypothesis or theory. When given two equally valid explanations for a phenomenon, one should embrace the less complicated formulation. The principle is often expressed in Latin as the *lex parsimoniae* (law of succinctness). For both of the pathologist studies, Take The Best achieved equally good predictions as the more complex competitors Franklin's rule and Dawes' rule did with a more economical explanation in terms of information used. While in case of Take The Best, for both samples nearly 100%⁵⁰ of the choices were predicted by using at most two pieces of information, Franklin's rule as well as Dawes' rule used all the five pieces of information to achieve the same performance.

⁵⁰ One choice of a US participant was explained by using three cues.

Therefore, I assume it to be reasonable to accept Take The Best as the best descriptor of pathologists' decision making with respect to a task such as that presented here.

Another interesting finding was that while German pathologists coming from university clinics opted more often for having nonmicroscopic tests than their nonuniversity-side counterparts did, for the respective US pathologists groups such a difference could not be found. This result indicates that within the USA nonmicroscopic test use, which usually goes along with higher costs, is no longer a privilege of university clinics. This might just mirror the increased desire and order behavior of tests of US oncologists documented by the main study results (see Chapter III) as well as reported by the US pathologists during the pilot study (see Pilot Study: *Results*) compared to the German counterparts. All pathologists described themselves as providing a service to the oncologists, in accordance with supply and demand. That no difference in choice behavior between US University and nonuniversity pathologists was found, however, while it was for the German ones, might reflect the effects of already well-established guidelines within the USA. Guidelines provide a common sense of what the current gold standard is to the US pathologists and even if pathologists were not to behave always in line with guidelines to a certain extent they will at least be influenced by this structured idea, which makes homogeneous decision-making amongst them more likely. This is not to say that German pathologists would not care about having an expert recommendation – when just taking a look at the most import cue of the Take The Best heuristic, it was for both the US and German samples the recommendation given by experts. However, the point is that guidelines within the German pathologist society are under development, but not established yet, and thus since there are not such clear prescriptions yet it is more likely that decision behavior is more heterogeneous.

Another point worth noting is that analytic specificity was not among the most important cues—or to put it more straightforwardly—was ranked on the last position. Although analytic specificity is not to be confused with diagnostic specificity and sensitivity (see *Implications for the Main Study*) it is still a measure of accuracy of a test. To find that it is neglected is something I view with some concern. One might argue that diagnostic sensitivity and specificity had mattered more to them, but if it were that far more than only two out of 93 participants within the German sample and 6 out of 108 within the US sample should had requested it within the main study, then. In a personal email a pathologist explained to the researcher that even to them figures of (diagnostic) sensitivity and specificity are rarely

available. Results of the pilot study pointed towards identical findings where diagnostic sensitivity and specificity were hardly mentioned. But analytic specificity was, and so it was considered as another measure of accuracy of a test performance to have a remarkable impact on choice, as also assumed by oncologists. In fact, at least in this study it did not. I can appreciate that pathologists might face the same problems and therefore aversions in working with probabilistic test information such as shown for other clinicians as well as it certainly is not a specific characteristic of a pathologist's brain to easily understand counterintuitive Bayes theorem. But this is exactly where it begins to worry me. Who is concerned about diagnostic sensitivity and specificity and who is able to process/calculate and understand these measures?

Unquestionably, authorities in charge of setting up guidelines should have such kind of knowledge and they surely do. Under the given circumstances of a clinician's working day such as the heavy workload, the time pressure, and expected difficulties in understanding probabilistic information it might be rather smart to rely on guidelines than to make up their own mind. However, within Germany such are mainly not established yet for pathologists and in addition, even in case they are presented with guidelines there might not always be scientifically objective and unbiased tools for guiding clinicians to the best standards of care as it were desirable. By just taking a look at recommendation by guidelines given on certain screening tests, one must sometimes wonder what the well-founded scientific reason is behind such recommendations. For instance in case of prostate cancer screening, studies have shown no actual mortality reduction by participating in such a screening program. The benefit of participation is marginal at most, while beside financial expenses to the health system the cost it could entail on patients ranges from psycho-emotional distress to serious bodily harm due to false test results and detection of benign tumors. And while for instance the U.S. Preventive Service Task Force concludes that there is insufficient evidence to either recommend for or against doing prostate cancer screening, others recommend in favor of such screening. That guideline recommendations do not always reflect a rational assessment of benefit and harm prove several drug regulatory decisions such as done on e.g. Kava⁵¹ as well (Greenhalgh, Kostopoulou, & Harries, 2004). It is not to assume that this will change with respect to the evaluation of pharmacodiagnostic tests as is certainly highlighted by introduction of the

⁵¹ Kava, a herbal anxiolytic, was widely used for centuries and has recently been shown to be hepatotoxic. A ban on it in North America has never fully implemented because of an indigenous minority view of this product as 'natural', and therefore safe.

Hercept-test, which received its approval just by the pressure of influential breast cancer patient groups in the USA (see Chapter II). Examples such as these make it apparent that guidelines always represent a compilation of certain interests and it might most of the times remain rather unclear to clinicians, which ‘politics’ are sold within the respective one. In that, there are seemingly some more open issues with respect to medical guidelines beside the one discovered already in Chapter III (see *Summary – Discussion*), which is not say that they are disregarded as a helpful decision aid for clinicians. If the underlying main interest is patients’ health they certainly are. Nevertheless, I prefer the vision of clinicians including pathologists being able to work properly with probabilistic information and therefore not having to rely solely on what guidelines or other experts regard as the best to do, but being able to draw their own conclusions. I proposed already within Chapter III (see *Summary – Discussion*) a simple natural frequency tree that can help to easily learn and communicate risk in a more effective way. In this sense, I would like to hold it with George W. Peabody’s (1922; cited from Casscells et al., 1978) statement and at the same time close this paragraph with: „*Good medicine does not consist in the indiscriminate application of laboratory examination to a patient, but rather in having so clear a comprehension of the probabilities of a case as to know what tests may be of value...it should be the duty of every hospital to see that no house officer receives his diploma unless he has demonstrated...a knowledge of how to use the results in the study of his patient.*“

In the upcoming Chapter VI methodological issues will be discussed.

5 The Idea of Shared Decision Making—Or: Does It Always Need Two to Tango?

“...the people whose preferences count are patients, because they are the ones who will have to live (or die) with the outcomes.”

D. M. Eddy

When investigating decision making in medicine, the times are gone when research only had to focus on actions taken by physicians, such as our Dr. Plaster. Although a certain amount of people still tend to regard solely him, or one of his colleagues in charge, of making proper medical decisions, attitudes about, and therefore models of, medical decision making have shifted in recent years from emphasizing the paternalistic responsibility of doctors toward increasingly advocating shared decision making as an ideal model of treatment decision making (e.g., Veatch, 1972; Brody, 1980; Quill, 1983; Brock & Wartman, 1990; Emanuel & Emanuel, 1992; Deber, 1994; Charles, Gafni, & Whelan, 1997, 1999). Several authors have described different ideas of the physician-patient relationship in different models (e.g., Emanuel & Emanuel, 1992; Charles et al., 1997).

The main purpose of this chapter is not to give a comprehensive overview about different archetypes of various models, however, but rather to introduce the main points of what role Dr. Plaster's patients are considered to play in the process of medical decision making, and which one they actually might desire to play, specifically when diagnosed with cancer. Therefore, only two main models, namely, the paternalistic and shared decision-making model, are briefly introduced for the sake of comprehension. The paternalistic model refers to a Dr. Plaster who ensures that his patient receives interventions that best promote their health and well-being, but which may not necessarily meet the patient's preference of care (Butow & Tattersall, 2005). The flow of information is one route from him to the patient, the information provided is of medical nature only, and the amount of information does not usually go beyond the margins of what is legally required. In contrast, shared decision making describes a partnership between Dr. Plaster and the patient, in which each contributes equally to the decision about treatment or care. Here, the information flow goes both ways, from the

doctor to the patient and vice versa. The exchanged information is not merely medical but also of personal nature, and the amount covers all relevant topics needed for the decision. Although these words depict the general idea of this concept quite well, it is necessary to emphasize that, due to the rather loose definition, today there is still confusion about which behavior falls into the boundaries of this concept and which does not. It therefore remains rather unclear to our Dr. Plaster how he could meet the demand of shared decision making properly.

Prior to the 1980s, the most prevalent approach to treatment decision making was the paternalistic one, with a Dr. Plaster assuming the dominant role and his patients in a minor position of just listening and following what they were told. This deference to professional authority was underlying a number of assumptions. One assumption was certainly that, for most diseases, only a single treatment existed and, therefore, there was not much to argue about. In cases where there was more than one treatment, however, clearly only physicians were regarded as knowing the best one, and able to consistently and appropriately apply information for selecting those due to their expertise, experience, and subsequent ability of valid clinical judgment. Because on top of this physicians were related to the professional concern of the welfare of their patients, they had a legitimacy in each treatment decision—a legitimacy further supported by professional codes of ethics which bound physicians to act in the best interest of their patients (Lomas & Contandriopoulous, 1994; Charles et al., 1999). All of these assumptions led both physicians and patients to accept the dominant role for physicians in treatment decision making. Status differences in terms of education and income surely powered the differentials in medical encounters.

Over the last decades, the credibility of the above assumptions began to be more and more questioned, and from there on an increased and fairly new interest in something called shared decision making started to infiltrate the opinion of what a good medical decision-making process would be. One reason that triggered the development was, without doubt, an increase in the number of diseases for which there was now more than only one treatment available. A more complex decisional context evolved where different treatments had different types of outcome in terms of benefit and risk. This rise in outcome possibilities offered an apparent choice, which might have fostered the courage to notice that it is the patient rather than Dr. Plaster who would have to live with the consequences of treatment outcomes. Subsequently, the view that it is the physician who was in the best position to evaluate and weigh these outcomes was increasingly challenged (e.g., Eddy, 1990; Levine, Gafni,

Markham, & MacFarlane, 1992). Around the same time, research focusing on the quality of medical care began to spotlight the effectiveness and appropriateness of the wide range of care delivered by physicians (Roos, 1984; Berwick, 1989; Lomas, 1990; Wennberg, 1990). There was increasing and consistent evidence that, for instance, physicians' procedures for the same disease often varied considerably across different small geographic areas and that these variations did not seem to be related to differences in the health status of the respective populations (Ross, 1984; Chassin, et al., 1987; Chassin, Brook, & Park, 1986; Leape et al., 1993; Iscoe et al., 1994; Charles et al., 1999). Variations in treatment patterns were even found for diseases for which clinical guidelines had been developed on best practices (Lomas et al., 1989; see also Chapter II). Patient preferences may have accounted for some of this variation, but the data also suggested that either some physicians were unaware of recommended best practices for the treatment of a specific disease or that they were aware of the respective guideline, but simply were not implementing it. In addition, more and more evidence was gathered regarding physicians' difficulties in judging correctly probabilistic outcomes (see Chapter II). Another health policy issue focusing attention on the physicians' performance definitely was concerns about rising health costs in western industrial countries, such as in the United States (Katz, Charles, Lomas, & Welch, 1998; for tests examples, see Cassells, 1978). The combination of cost and quality concerns resulted in recommendations to make physicians more explicitly accountable to patients and the public (Charles et al., 1999), and finally led to the consequence of patients' involvement in their medical decisions.

5.1 Is to Tango Desired in the Field of Cancer?

While shared decision making is now regarded as the gold standard, and Dr. Plaster is more or less urged to ensure that this occurs in his practice, there are many unanswered questions about its meaning and application (Butow & Tattersall, 2005). For certain diseases, such as cancer, which are both potentially life threatening and widely prevalent, treatment decision making is quite complex in this setting. Despite the publication of several quantitative studies (e.g., Fallowfield & Jenkins, 1999; Meredith et al., 1996; National Cancer Alliance, 1996; Jenkins, Fallowfield, & Saul, 2001), physicians and health professionals remain concerned about the amount and type of information to give to a patient with cancer. Decisions often involve weighing up uncertain benefits against uncertain side effects. For example, the risk of cancer recurrence after surgical treatment may be reduced from 40% to

20% by administering additional chemotherapy, but there is no way of knowing at the time of decision making whether the patient is still within the 40% at risk or within the 60% who are cured by surgery alone. Furthermore, different treatments may have the same benefits, but differ in their side effects, that is, while some patients receiving chemotherapy suffer severe side effects, others may have a fairly mild experience, which also cannot be easily predicted upfront. Even though pharmacodiagnostic tests would focus on exactly this problem, they come with their own probabilistic uncertainties (Chapter I), which are surely as hard to handle as the current ones and, therefore, are not the ultimate solution for helping to clear up the fog around the benefit-risk issue in cancer care.

Most importantly, in this respect, is certainly that the outcomes of the decisions are major, involving life and death. Therefore, decisions in the field of cancer may differ from medical decisions in many other medical fields. Although the majority of cancer patients in the western world seem to desire full information about their disease (Cassileth, Zupkis, Sutton-Smith, & March, 1980; Degner, Kristjanson, & Bowman, 1997; Jenkins et al., 2001; Davidson, Brundage, & Feldman-Stewart, 1999), there is still considerable variation in the extent to which patients wish to participate in the decision making (Degner et al., 1997, Butow et al., 2004). Patients are often traumatized by the cancer diagnosis or news about cancer recurrence and are usually in a highly emotional state, which might impair cognitive abilities needed for making a well-defined decision. Several qualitative studies have found that patients report being overwhelmed with emotion at the time of making such treatment decisions and do not feel able to make such a sensible decision (Charles, Redko, Whelan, Gafni, & Reyno, 1998; McVea, Minier, & Johnson Palensky, 2001; Lam, Fielding, Chan, Chow, & Or, 2005), and prefer to defer the treatment decision to their physician (e.g., Davison, Degner, & Morgan, 1995; Degner & Sloan, 1992; Sutherland, Llewellyn-Thomas, Lockwood, Trichtler, & Till, 1992). Apart from the fact that patients frequently feel that they lack the expertise to make a decision, basic details of diagnosis, prognosis, and treatment information that were delivered are commonly reported to be misunderstood amongst cancer patients (Gattellari, Butow, Tattersall, Dunn, & McLeod, 1999), which might be a function of both problems in adequately communicating probabilistic information on the side of the physician as well as impaired cognitive abilities due to the highly emotional state on the side of the patient. Cancer patients often look to their oncologists to “save their life,” seek from them reassurance and hope, and feel a strong need to trust their skills and expertise (Butow & Tattersall, 2005). In a recent

study by Henman, Butow, Brown, Boyle, and Tattersall (2002), 20 women with heterogeneous cancer were interviewed about the components of optimal decision making. This study found that if women felt confident that the doctor cared for, understood, and respected them, and that they could trust and have confidence in the doctor, they were pleased to accept the doctor's recommendation without engaging in their own decision making. Even when these patients reported a desire for collaborative decision making, they relied heavily on their doctor's opinion, and sought rather to understand the rationale behind this recommendation than to make the decision by themselves. In a quantitative study, Salkeld, Solomon, and Butow (2004) found, by asking 175 male and female colorectal cancer patients to rate the importance of various aspects of the treatment decision-making process, that aspects of trust in the surgeon were rated by almost 100% of the participants as being important, while other aspects scored 66% and lower. Indeed, whilst the outcome of treatment, such as survival, side effects, and quality of life, are very important to patients, the trust in the oncologist seems to be of utmost importance to them.

Such findings might be a reason why several authors have suggested that shared decision making is not always realistic or desired by patients (Lam et al., 2005). A variety of factors may contribute to the extent of the cancer patients' desire for deferring the decision to their oncologist, but the already above-mentioned ones, such as the high degree of complexity and uncertainty underlying treatment decisions in the field of oncology, surely are the main drivers of patients' application of such a kind of expert-opinion heuristic. We already noted the difficulties that physicians have when handling excessive amounts of medical information (Chapter II). However, cancer patients, not only novice but also in an especially elevated emotional state, are even more likely to find the sheer amount of available data overwhelming. This kind of "cognitive overload" may make patients rather less than more able and willing to contribute to the decision-making process. But could the same picture be drawn for the patients' decision on a pharmacodiagnostic test? From Fall 2004 to Spring 2005 I examined, with an explorative pilot study in Germany and the USA, what role patients play in the decision-making process on these tests, which role they might desire to play, and what cues would drive their decision on wanting to see such tests apply to their own treatment decisions by the oncologist.

5.2 Pilot Study

“The value of life lies not in the length of days but in the use you make of them.”

Montaigne

5.2.1 Design

The pilot study was aimed at gaining knowledge of the role of cancer patients regarding their involvement in test decisions, identifying relevant determinants of their decision-making, finding out more about the environmental conditions and constraints of the decisions, as well as about the decision-making strategies that patients would apply. For this reason, semistructured interviews with an open answer format were applied. Again, this investigation was conducted by gathering a list of decision-making relevant cues for patients as well as to generate adequate hypotheses, as no research has addressed this question hitherto.

The first question of the interview determined whether or not the participant had ever experienced making a decision on a pharmacodiagnostic test.⁵² The patients who had apparently neither experienced a choice on such types of tests, or had even heard about it, continued immediately with the questions concerning the decisions on any other type of testing, such as needle biopsy in breast cancer⁵³ as well as on their treatments that had been taken in the past. This procedure was chosen, as it was assumed that patients might apply comparable decision cues for any type of test as well as for treatments, as they would for pharmacodiagnostic tests. The assumption that cues applied for any type of testing and treatment decision were similar was checked for each participant while the interview took place, that is, it was monitored whether or not participants mentioned similar cues. If participants were never faced with a decision on any test, they were asked to imagine a decision for a hypothetical testing situation, from which answers were taken as a check.

⁵² When the patients' interviews took place in Fall 2004 to Spring 2005, some target tests (see Chapter I) had already been launched.

⁵³ Part of the most diagnosis processes for several cancer types are invasive procedures, such as the examination of the spinal cord in the case of leukemia or biopsy performed on the affected breast in the case of breast cancer.

Table 5.1: Catalogue of the Interview Questions for Patients

Topic	Questions
Demographic data of investigated person	Age Gender: Type of cancer Stage of cancer Social and financial status/profession Received treatment
1. Questions on experience	<p>1.1 Do you know if your doctor used a test to better identify the treatment that would help you best?</p> <p>1.1.1 If so: Can you remember what the main purpose of the test was?</p> <p>1.1.2 Did your doctor explain to you what the use of the test is or which results are possible?</p> <p>1.1.3 Did you have the desire to actively take part in the decision on the test?</p> <p>1.1.3.1 If yes: Did you feel that your doctor was giving you the opportunity of doing so?</p> <p>1.1.3.2 To which extent did you end up participating in the decision?</p> <p>1.1.4. Did you feel that you had received all the information necessary for such a decision from your doctor?</p> <p>1.1.4.1 If so: Which information did you receive?</p> <p>1.1.4.2 If not: Which ones would you have liked to received?</p> <p>1.1.5 Did your doctor give you any information about the result of the test?</p> <p>1.1.6 Did you have any concerns regarding the test?</p> <p>1.2 When the decision on your last therapy was taken, what exactly happened? Please describe the situation as exactly as possible?</p> <p>1.2.1 Did you have the desire to actively take part in the decision on your treatment?</p> <p>1.2.1.1 If yes: Did you feel that your doctor was giving you the opportunity of doing so?</p> <p>1.2.1.2 To which extent did you end up participating in the decision?</p> <p>1.2.2. Did you feel that you had received all information necessary for such a decision from your doctor?</p> <p>1.2.2.1 If so: Which information did you receive?</p> <p>1.2.2.2 If not: Which ones would you have liked to receive?</p> <p>1.2.3 Did you use any additional sources of information? Which ones?</p>
2. Questions concerning responder and prognostic tests	<p>Providing an explanation of responder and prognostic tests:</p> <p>2.1 Do you expect that such future tests will be beneficial to cancer therapy?</p> <p>2.2 Could you also imagine potential downsides?</p> <p>2.2.1 If yes: Which ones?</p> <p>2.3 Can you imagine circumstances under which you would not allow such a test to be carried out on you?</p> <p>2.4 Can you think of any features of the test/the manner in which the test is carried out that could lead you to declining the application of such a test on you?</p> <p>2.5 If the doctor was to carry out such a test on you, do you think they should inform you?</p> <p>2.5.1 If yes: What should they tell you about the test?</p> <p>2.6 Would you be ready to pay for such a type of test if your health insurance would not cover the costs?</p> <p>2.6.1 If so: What would it depend on?</p>

Note: Not every single participant was necessarily asked each of the questions. It did depend on the answers given for each question what questions were asked next.

Further questions addressed the extent of their involvement in the most recent test and treatment decisions and their general desire for participating as well as their seeking behavior for information apart from those provided by their oncologists.

Finally, the concept of predictive and prognostic tests was explained to patients, and they were asked to think of information that they would desire for having such a test applied, and any reason that would make them opt for, or against, such a test.

In order to learn about any kind of intragroup differences that might affect answers given, demographic questions as well as cancer-related questions were asked at the beginning of the interviews. Table 5.1 summarizes each of the possible questions.

5.2.2 Procedure

In order to acquire access to patients, German pilot study oncologists were asked for their support by making contact with their patients. Typically, the oncologists asked some of their patients if they were willing to participate in this study. When they agreed to participate, contact data, such as a telephone number, was revealed to the investigator. In the USA, two cancer patient organizations, namely the *Gilda's Club* and the *Komen Breast Cancer Foundation* in Seattle, were asked to help to find some members who would be prepared to participate. Only patients receiving treatment were included in the pilot study to ensure that each participant had already experienced at least one decision about treatment. However, no limitation was set to the type of cancer they had. Each patient who agreed to participate was contacted upfront, and an appointment made. Participants were usually interviewed either at home, at the premises of the respective patient organization, or at a café in the vicinity of their home. All participants were explicitly informed that their answers would remain completely anonymous and confidential. The interviews ranged from 30 minutes to 2 hours (Germany, mean: 95 minutes; USA, mean: 78 minutes). Patients received no compensation for their participation.

5.2.3 Participants

Ten German cancer patients, six women and four men from Berlin, Potsdam, as well as from Leipzig, and seven US patients, five women and two men, all from Seattle (WA) were interviewed. Table 5.2 summarizes the characteristics for the German and the US sample.

Again, due to time constraints, not as many patients were interviewed for the US pilot study as for the German study.

The German study ($N = 10$) comprised mainly four breast cancer patients, four leukemia patients, and one colon cancer patient, as well as one non-Hodgkin lymphoma patient, that is, five patients were treated with a solid tumor and five with hematological ones. The age of the German sample ranged from 22 to 73 years (mean: 53, SD : 17.04). Of the ten patients interviewed, six were diagnosed as being at an adjuvant stage, and four at an already metastatic stage. Three patients had received targeted therapy.

For the US sample ($N = 7$), which ranged in age from 31 to 73 years (mean: 51, SD : 12.50), the majority of patients suffered from breast cancer ($N = 4$), one patient from ovarian cancer, one from brain carcinoma, and one patient from non-Hodgkin lymphoma. Four patients were diagnosed at the adjuvant stage of their disease, while the remaining three patients were at the metastatic stage. Out of the seven interviewed patients, three were being treated with target therapy. No significant differences were found between the two subgroups, except from the type of cancer and the location where the treatment was received. This, however, was not found to be influential for the final results derived from the pilot study.

Table 5.2: Description of the Seventeen Participating Pilot Study Patients

Demography	Germany ($N = 10$)	USA ($N = 7$)
1. Age	22–73 years (mean: 53; SD : 17.04)	31–73 years (mean: 51; SD : 12.50)
2. Gender	Male: $N = 3$ Female: $N = 7$	Male: $N = 2$ Female: $N = 5$
3. Education	Academic: $N = 5$ Nonacademic: $N = 5$	Academic: $N = 5$ Nonacademic: $N = 2$
4. Type of cancer	Breast cancer: $N = 4$ Colon cancer: $N = 1$ Leukemia: $N = 4$ Non-Hodgkin Lymphoma: $N = 1$	Breast cancer: $N = 4$ Ovarian cancer: $N = 1$ Non-Hodgkin Lymphoma: $N = 1$ Brain cancer: $N = 1$
5. Cancer stage	Adjuvant: $N = 6$ Palliative (metastatic): $N = 4$	Adjuvant: $N = 4$ Palliative (metastatic): $N = 3$
6. Type of treatment^a	Chemo therapy: $N = 8$ Target therapy: $N = 3$ Hormonal therapy: $N = 3$ Radiation: $N = 4$	Chemo therapy: $N = 5$ Target therapy: $N = 3$ Hormonal therapy: $N = 2$ Radiation: $N = 3$
7. Where treated	University/research clinic: $N = 5$	University/research clinic: $N = 4$

Hospital: N = 2	Hospital: N = 3
Private practice: N = 3	Private practice: N = 0

^aNumbers do not add up to the number of participants, as several patients received more than one treatment.

5.2.4 Analytical Procedure

The analytic procedure of the 17 patient interviews was identical to that for the oncologist (Chapter III) and pathologist (Chapter IV) sample. All interviews were transcribed; each of the interview protocols was studied and searched for cues mentioned to be important for the decision. Cues mentioned by at least 50% of the participants were extracted for each of the respective groups for the main study. Results of both the German and the US subsample were again compared with each other.

5.2.5 Results

“Well, he thought it would probably be the best for me and I agreed....”

A pilot study patient

Almost none of the patients (N = 14 – for both) reported having any experience or familiarity with pharmacodiagnostic tests.⁵⁴ Of the three patients coming from the USA who had knowledge about such tests, one knew about it by profession while the other two patients were acquainted with such tests as the treatment they finally received⁵⁵ was based on a result of a subgroup of such tests (target tests). Asked about how the decision on the test were taken, both patients reported that they had been informed about the implementation of the test at the same time as what the decision on their management of care is based on was explained to them. Both patients, however, said they were satisfied with how this decision was taken by their oncologists.

Do patients face a decision on a test at all? All patients of both groups reported to have experienced some type of testing before the treatment administration was started. Patients were normally only informed about a test upfront when it was associated with an invasive procedure. For all noninvasive tests, information about a test taken was only found to be

⁵⁴ Note that participants were not asked literally for a pharmacodiagnostic test (see Table 5.1).

⁵⁵ Herceptin[®] in the case of a breast cancer patient at the palliative stage; Erbitux[®] in the case of a colon cancer patient at the palliative stage.

provided in the USA and only in cases where it had effected the management of care. In the case of invasive procedures, patients usually only received an explanation of how the test was going to be implemented and whether taking the test could cause any bodily harm. That a test can provide false results was not mentioned to the patients, and, therefore, was an unfamiliar fact to each of them. When asked what led them to agree to an invasive test, all patients reported to have based their decision on the recommendation of the oncologist who hinted at the importance of the test for a precise diagnosis. Patients were also asked if they had tried to obtain any further information about the test, which was usually not the case. All interviewees pointed out that the explanations given by the oncologist had been sufficient for them to make the decision to have the invasive testing.

Getting involved in all that? In the pilot study, little involvement in any medical decision was noticed by the reports given, a fact that proved to be true especially for the German sample. Although there were differences in “activity” between the German and US patients, that is, US patients reported networking activities within patient groups (N = 6) and to have requested a second opinion by another oncologist (N = 4), nearly all of the interviewees (N = 16) had based the decisions for their treatment or testing in the past on what the oncologist had recommended. Especially women reported feeling emotionally too overwhelmed by the fear of death, and the uncertainty of what the diagnosis might mean for their life, to take any action. Apart from three patients, none of these patients had felt patronized by their clinicians, but even felt involved and informed. For the three who had felt patronized, two eventually followed the recommendation of another oncologist, while the denial of the recommendation of the former oncologist was mainly driven by missed sympathy and warmth. Only one female patient did not follow her doctor’s recommendation at all. Inspired by a healer’s book that highlighted the evil of chemotherapy, she decided completely against receiving any treatment at the outset. Four women from the German sample as well as three women from the US sample reported that their oncologist had to take extra trouble to convince them of receiving treatment, which was supposed to cause severe side effects. If a test was available for finding out their eligibility for the treatment in question, these patients stressed that the chance was high for them to decide on having the test without the generally important recommendation of their doctors.

Asked whether additional sources of information were used, both samples quoted the internet and disease-specific books. For the US sample, patient groups were indicated as an extra source of information. Information that patients sought out was related to the current standard of treatment, enabled patients to learn more about survival rates, informed them of where to receive emotional support, and where to find a suitable patient group. In this study at least, patients stated that when they gained any information on promising therapies or even tests through the above sources of information, they discussed them with their oncologist and normally follow their recommendation.

When interviewees were asked questions concerning responder and prognostic tests (see Table 5.1) after introducing these tests to patients, still most (N = 15) felt not competent enough to answer the questions in the second part of the interview. A large part believed that such tests would be definitely helpful to have, especially since most patients felt concerned about the severe side effects that most cancer treatments cause, but that, ultimately, they doubted being able to estimate the tests' ultimate value, and would thus leave the decision to their oncologist. If the oncologist deemed the tests useful, none of the interviewees could picture a reason against their usage. With respect to the type of sample needed to perform the test on (e.g., tissue), all patients stated that they would be ready to accept a high degree of invasiveness for maximum therapy benefit and minimum side effects. Also, all US patients and most of the German patients (N = 7) were willing to pay for such a test should their oncologist deem it significant for the treatment decision, but the price would definitely determine the limit.

Table 5.3: *Cues Important for Patients' Choice on Applying a Pharmacodiagnostic Test Presented by Number of Mentioning*

<i>Cues</i>	<i>Germany (N = 10)</i>	<i>USA (N = 7)</i>
<i>Recommendation given by the oncologist</i>	N = 10	N = 7
<i>Potential cost of the test</i>	N = 10	N = 5
<i>Severity of the side effects of the treatment</i>	N = 5	N = 3

With respect to patients' applied decision-making strategy, the all-striking rule was to follow the recommendation of their oncologists. Nevertheless, the relationship to the doctor,

self-coverage of test costs, and over-duly concerns about experiencing severe side effects of the treatment could negatively impact the influence of the oncologist's recommendation. Even though such minor groups exist, for the majority of cancer patients within this study, the recommendation of their doctors was the one single reason that influenced their decision making, and they were seemingly satisfied with this. Table 5.3 lists the important cues of patients' decision making regarding pharmacodiagnostic tests.

5.3 *Implication for the Main Study: Is There Anything to Tango Regarding Tests?*

The recent call for greater patient participation is surely based on the assumption that patients desire and benefit from having a more involved position in the medical decisions related to them. However, apart from the fact that participants seem to confound participation with information,⁵⁶ in this pilot study, it was in accordance to other studies' findings in the field of cancer that most of the patients wanted their clinicians to take a primary role in the decision making. After diagnosis, most patients found themselves in a highly emotional state and felt any disease-related decision to be specifically challenging. Patients simply felt that they were not able to make such a sensible decision involving life or death, and especially not regarding any elaborate tests. Having their clinicians make the decision for them was, for the most part, seen as relieving and also not deviant from what they would call normality in terms of making medical decisions. The first hypothesis to test was, therefore, whether patients apply the simple rule of thumb of following the recommendation of the doctor when having to decide on a pharmacodiagnostic test:

H1a: If conducting a pharmacodiagnostic test is recommended by a patient's oncologist, the patient would significantly opt more for having the test than if it is not recommended.

It was noticed, however, that the intensity of the oncologist-patient relationship could be either a promoter or an obstacle for applying this rule of thumb. Patients who missed warmth and trust in the oncologists-patient relationship and experienced the relationship as being reserved seemed less likely to follow their oncologist's recommendation.

H1b: When patients do not experience their relationship to the oncologist as trustful and warm then they follow the recommendation of their oncologist less often than those who did.

⁵⁶ Based on participants reports, it was noticed that most patients felt to have participated in the decision if the oncologist had informed them of what they think might be the best thing to do. Furthermore, patients mainly based their evaluation on whether they were properly informed and involved in the type of relationship they had with their doctors. If this was characterized by warmth and based on trust, patients were happy to follow the recommendation of the oncologist, together with the impression of having been a part of the decision. This, however, is different from what the concept of shared decision making demands.

Although all US patients and most of the German patients expressed their willingness to pay for such a test, in cases where it is not covered by health insurance, its price could definitely become an eliminating criterion. Especially for German patients, who are still quite unfamiliar with paying for health services themselves, they felt uncomfortable with this idea and deemed paying for a test, to a higher extent, an insurmountable obstacle, compared to their American counterparts. It is thus assumed that there is a certain level of test price that will make people decide against the test even if it is recommended, and although the recommendation is expected to be still a paramount criterion for the decision.

H2a: If patients were to pay for the test, it would diminish the impact of a recommendation for taking the test given by the oncologist.

H2b: This effect is stronger for the German sample than it is for the US sample.

Another factor recognized as being able to bring an oncologist's recommendation to sway was particularly strong concerns of a patient regarding the side effects of the treatment. If a patient was particularly concerned about experiencing such severe side effects from the considered treatment, they were expected to be more likely not to follow a recommendation against such a test, but to have the test in order to make sure that they were eligible for the treatment in question. Although not often, also pilot study oncologists reported such situations occurring in their practice when patients had heard about new ostensibly promising therapies or, more recently, even tests. So again, the "rule of thumb" to base the decision solely on the recommendation given by the oncologists would be challenged, in this case by the severity of the treatment side effects, which would result in an especially concerned subgroup of patients opting for the test and deciding against an "against-recommendation."

H3: Particularly concerned patients less often follow an oncologist's recommendation against having the test than not particularly concerned patients if the severity of the treatment side effects is high.

The rationale of the patients' pilot study was to learn and understand how patients would proceed when they were offered a decision on having a pharmacodiagnostic test for

deciding on their management of care, and, furthermore, to be enabled to generate meaningful hypotheses. The rationale of the main study is to test whether these hypotheses would withstand statistical testing in order to learn about what mechanism and cues patients finally apply when facing the decision regarding a pharmacodiagnostic test, and if assumed intragroup differences do indeed exist. Again, the method chosen for the main study to elicit cues of relevance and decision-making processes is discrete choice experiments. Within the next paragraphs I will give detailed information about the design, participants, procedure, and lastly, the results of the main study.

5.4 Main Study

5.4.1 Introduction

Identically to the ambitions regarding the main studies of oncologists and pathologists, I was also curious, with respect to the patients, whether or not their decision making would be described and predicted equally well by a fast and frugal (noncompensatory) model than it would by a compensatory integrative one. Furthermore, I wondered again, if there were any intercultural differences that would show a more substantial and impressive effect than those found in the pilot study when conducting a study on a larger scale. For the sake of all these desires, I also set up a main study for the patients, which will be described in the following.

During the pilot study, it was found that cancer patients mainly use a simple “rule of thumb,” namely, to follow the recommendation of their oncologist when facing a medical decision. While cues were reported having the potential to diminish or even completely suspend that heuristic, it was expected that patients’ decision making is at least as well-described by a fast and frugal model as it is by a compensatory integration model, due to the patients’ emotional state after diagnosis and their feeling of the lack of medical expertise. It is furthermore hypothesized that a fast and frugal model is also able to at least as well-predict unknown choices as a compensatory integration model.

H4a: A fast and frugal model will equally well fit patients’ decision-making data as a compensatory integration model.

H4b: A fast and frugal model will equally well predict unknown patients’ choices as a compensatory integration model.

5.4.2 *Design*

The same methodological approach, which was applied for the oncologists' main study (see Chapter III) was also employed for the patients' main study, that is, fairly comparable procedural steps were taken regarding the design. Therefore, only the steps that were different from those in the oncologist study will be reported in the following.

The following cues were found to be, to some extent, decisive for the decision: Test recommendation by the oncologist, the cost of the test, the severity of the side effects of the treatment the test focuses on, as well as the relationship to the oncologist. Since during the pilot study it was experienced, that patients are rather unfamiliar with a choice on tests, it was sought to keep the cases as simple as possible in order not to induce noncompensatory decision making by overchallenging the participants. Therefore, the first three cues were chosen to be manipulated, the last cue was asked for in the demographic question section (see Table 5.5) in order to investigate its impact on choice. Due to the small number of cues, I was able to apply a full factorial design. The design ensured orthogonality, furthermore, all main effects were verifiable and intercorrelations between the cues were kept down to zero.

Out of the three cues that were used to construct the cases, two were set at three levels and one cue at two levels, which finally led to 18 cases. With consideration of patients' morbid health state, it was opted for presenting the questionnaire again randomly split in a version A and B, which resulted in nine cases per questionnaire version. The cue levels were specified in a consent process with two patients and one oncologist to ensure that cases would reflect what could happen in reality. Each of the cue values was distributed equally among the set of the 18 cases as well as among each version of the questionnaire. The three cues, their levels, and the distribution of their values are shown in Table 5.4.

To outline the cover story, a general description of a fairly common decision situation in oncology was delivered to the patients, stating that they should imagine that their doctor (oncologist) wants to add a further therapy X to the therapy they were already receiving. They were furthermore told that therapy X is a standard therapy, from which 5 out of 100 patients benefit. Then, they were asked to imagine that therapy X would come with a test to find out who would benefit from the therapy. However, since having this test is not as yet required, it was up to the patient to decide whether or not they wanted the doctor to apply it. To make sure that main study participants would really understand what impact having the test would have

on any further treatment decisions, it was further explained to them that if they decided on receiving the test, the administration of therapy X would be based on the test's result. If they decided against having the test, they would still receive therapy X insofar as they wanted to, as it remains the standard treatment for their cancer. As one cue value of *recommendation by oncologist* offered to the patients was the value *nonrecommendation of oncologist* regarding the test, it was felt that patients might wonder why a doctor should do so. In reality, they could have asked, but not when using an online or paper questionnaire. Therefore, a short explanation on why doctors may give a recommendation against having a test was added.⁵⁷

Before starting to present the cases, participants were told that there are neither right nor wrong answers regarding the upcoming choices, and were kindly asked to make their decisions within the cases just as they would in a real medical situation. Then the presentation of the nine cases of either version A or B of the questionnaire followed. For each case, patients were asked to give a yes/no response to whether or not they wanted their oncologist to apply the pharmacodiagnostic test under the described conditions of the cases. Since in reality patients furthermore have the opportunity to completely decide against receiving any treatment, that is, also without having any test, this was offered as another possible choice option. All participants, who chose this opportunity in the main study were excluded from the further analysis, as this study did not aim at understanding treatment decisions, but decisions on tests. That is, although this choice opportunity was visible to participants, for the sake of depicting the reality, this choice opportunity was not existent for this study.

Again, four questions, with regard to the perceived representativeness of the described situation and the offered cases, were presented to participants after cases were shown (see Chapter III: *Table 3.5*). In order to investigate whether differences in the doctor-patient relationship, in fear of experiencing side effects, in the degree of participation in the most recent decision,⁵⁸ or any other kind of common demographic values, might have an effect on the participant's choice, several questions addressing these issues were asked at the end of the questionnaire. The respective questions with their values are outlined in *Table 5.5*.

⁵⁷ "Please note that there can be various reasons for your doctor recommending or not recommending the test. The main reason, however, is that the quality of the test might not be good enough and might have a high probability of a false result. The false result might then also lead to the wrong decision on your treatment."

⁵⁸ I decided to ask for participation issues with respect to the last treatment decision. Based on the experience from the pilot study, I assumed not to derive any meaningful data with respect to this issue when asking the same for a test decision due to the current medical practice.

Table 5.4: *Manipulated Cues, Their Levels and Distribution for Version A and B of the Questionnaire*

<i>Cues</i>	<i>Level</i>	<i>Distribution Part A</i>	<i>Distribution Part B</i>
Severity of the side effects of the treatment ⁵⁹	(0) Therapy X causes moderate side effects in the majority of patients. 10 out of 100 patients may even have serious side effects.	(0) 5	(0) 4
	(1) Therapy X causes serious side effects in the majority of patients. 5 out of 100 patients may even have life-threatening side effects.	(1) 4	(1) 5
Cost of the test	(0) Covered by health insurance.	(0) 3	(0) 3
	(1) €/ \$ 300	(1) 3	(1) 3
	(2) €/ \$ 2,000	(2) 3	(2) 3
Test recommendation by oncologist	(0) Your doctor strongly advises against using the test, because s/he has doubts about the quality of the test.	(0) 3	(0) 3
	(1) Your doctor does not see any reason to use the test but also no reason not to use it, so the decision is left to you.	(1) 3	(1) 3
	(2) Your doctor strongly recommends using the test, because s/he thinks it is high quality and provides important information for use in your treatment.	(2) 3	(2) 3

The main study task was piloted on four German and three US participants of the pilot study, and minor adjustments on making more obvious within-choices, that the side effects are related to the treatment and not to the test, were carried out.

Table 5.5: *Questions and Their Values for Capturing Intragroup Differences and Demographics*

<i>Questions</i>	<i>Values</i>
D1: How old are you?	Open field
D2: What is your gender?	– Female – Male

⁵⁹ On the webpage, patients were provided with a clickable overview comprehensively explaining what the different side-effect levels mean.

D3: What is your highest level of education?	<ul style="list-style-type: none"> – University, bachelors, or graduate degree(s) – Associate degree, community college – High-school graduate – Some high school – None Open field
D4: Which kind of cancer did/do you have?	
D5: At what stage was the tumor at the time of the (last) diagnosis?	<ul style="list-style-type: none"> – Metastatic – Nonmetastatic
D6: What type of treatment have you received? (Please check all that apply.)	<ul style="list-style-type: none"> – Chemotherapy – Hormone therapy – Antibody/target therapy – Surgery – Radiation – Other
D7: When your doctor decided on your treatment, did you wish to participate in the decision?	<ul style="list-style-type: none"> – Yes, I did. (please continue with question D8) – No, I did not. (please continue with question D10) – I do not remember if I wanted to participate in the decision.
D8: Did you feel that your doctor gave you the opportunity to take part in the decision about your treatment?	<p>(please continue with question D10)</p> <ul style="list-style-type: none"> – Yes, they did. (please continue with question D9) – No, they did not. (please continue with question D10) – I do not remember. (please continue with question D10)
D9: To what extent were you eventually able to participate in the decision?	<ul style="list-style-type: none"> – My doctor gave me a recommendation; the final decision was up to me. – The decision was made jointly with the doctor. – I do not remember.
D10: Upon receiving your diagnosis, did you try to get more information on your disease and/or your treatment options?	<ul style="list-style-type: none"> – Yes. (please continue with question D11) – No. (please continue with question D12)
D11: Which sources did you turn to? (please check all that apply)	<ul style="list-style-type: none"> – Books/journals – Internet/media – Consulting other specialists (e.g., other oncologists, doctors) – Patient groups – Friends/family
D12: Before treatment, did you worry about the potential side effects of the therapy?	<ul style="list-style-type: none"> – Yes, I did worry. (please continue with question D13) – No, I was not very concerned. (please continue with question D14) – I do not remember whether or not I was concerned.
D13: How strong was your concern about side effects?	<p>(please continue with question D14)</p> <ul style="list-style-type: none"> – I was so concerned that I refused or almost refused the recommended treatment. – I was concerned, but would not have refused a drug recommended by my doctor. – I was concerned to minor extent, but this did not play any decisive role.
D14: Which of the following best describes your relationship to your doctor? (please check all that apply)	<ul style="list-style-type: none"> – The relationship with my doctor was trusting and positive. They have taken my worries and fears seriously. – The relationship with my doctor was neutral. They have not always taken my worries and fears seriously. – The relationship with my doctor was negative/cool/reserved.

They have hardly/never taken my worries and fears seriously.

D15: How much have you spent in the past Open field
for additional over-the-counter drugs?
(e.g., alternative treatment) (\$/€ average
per month)

5.4.3 Procedure

A detailed description of the decision situation in question and the nine related cases were presented on either a webpage or a paper version to the participants. Upfront, a short foreword was included, which introduced the study as well as the investigator, and highlighted and guaranteed respondents' anonymity. At the bottom of the webpage, questions considering the intragroup differences and demographic data as well as questions regarding the quality of the questionnaire were included. Either version A or version B was presented randomly to each of the participants, whereas only the cases presented, but not any other part of the questionnaire, such as the description of the medical situation or demographic questions, were varied. As the cover story outlined a fairly general decision situation, no restriction on the type of cancer existed, that is, all patients diagnosed with any type of cancer were regarded as eligible to respond to the online questionnaire.

In Germany, to announce the online investigation, several cancer patient-orientated online forums were approached, and asked whether they were prepared to provide a short description of the project along with a clickable link to the online study on their respective webpage. When they agreed, the relevant information was sent to them (see Appendix C). The following forums finally posted information about the project and the clickable link on their webpage: *Inkanet*, *Mammazone*, *Krebs-Kompass*, *Tumor Centre of Regensburg*⁶⁰ and the online newspaper "*Mensch & Krebs*." It was assumed, though, that having online access is not standard in the sample of patients, therefore, a paper version of the questionnaire was prepared alongside the online questionnaire⁶¹. The paper versions were completely identical to the webpage with respect to the content. To enable the distribution of the paper versions, again some pilot study oncologists and some hospitals were asked for their assistance. Three oncologists, two located at the university clinic Charité Mitte as well as in Leipzig and one at the doctor's own cancer focal point practice in Berlin agreed to ask their patients to volunteer in the main study. When patients agreed, the oncologists handed out the paper version of the

⁶⁰ Tumorzentrum Regensburg eV; Online-Zeitschrift "Mensch & Krebs"

⁶¹ Online version and paper version of the questionnaire were identical in content.

questionnaires to the patients. To work through the questionnaire, patients were allowed to take this home, and were requested to bring it along when their next appointment was due to take place. Further paper versions were personally distributed to cancer patients at the oncological day care unit of the clinical centre in Magdeburg by the investigator. After a short introduction of the investigator by the oncologists, the investigator gave a short explanation of the purpose of the study and asked for participation. When the patients agreed, the paper version of the questionnaire was handed out to them and collected after its completion, which was usually the case after 15 to 30 minutes. With respect to the response rate, in the case of the distributed paper version by the three oncologists and the investigator, almost each approached patient participated, which led to a final response rate of some 90%. For the online version, where a link to the study was provided on several cancer patient-related webpages, the response rate was indeterminable, but estimated to be quite low, given the noticed impact that the announcements finally had on that study.

For the US sample, I was fortunate to be able to revert to an already more established patient group system than in Germany. A fairly large cancer patient organization, namely Gilda's Club, which has branch offices all over the USA, was considered as being helpful for spreading information about the study. All of the 19 affiliations were contacted and asked for their support by distributing an email to their members,⁶² which provided patients with a short explanation of the investigator and her research, the request for participation, as well as a link to the respective webpage (Appendix C). The following affiliations finally provided their members with the email: *Gilda's Club*, in *Seattle*, *Southeastern Wisconsin*, *Quad Cities*, *Westchester*, *Detroit*, *Grand Rapids*, *Capital Region New York in Albany*, *New York City*, and *North Texas*. Furthermore, the *Susan G. Komen Breast Cancer Foundation* affiliate in Seattle provided a short description of the project together with a clickable link to the online study on its respective webpage to their patients. Although asked for, none of the Gilda's Club branches revealed a precise number of the email contacted members, apart from the Gilda's Club in Seattle, which estimated the number of some 1,000 contacted patients. Considering the number of responses to the questionnaire after the announcement of the study conducted by the Gilda's Club in Seattle had taken place, the response rate for this particular club was some 4%, which is rather low. However, such rather low response rates seem to be quite typical

⁶² They were explicitly asked not to reveal their member lists to the investigator to ensure the highest level of privacy for the participants.

when no personal engagement between the researcher and participants exists. Another issue for this low response rate might have been the task itself, which was intellectually challenging, in particular due to its unfamiliarity to patients. The final response rate was not estimatable, since, in cases where personal contact to the patient existed, nearly each patient participated, which was approximately 90%; however, in cases where no personal contact was involved, I saw a response rate of only some 4%. Furthermore, for most of the sources used, precise response numbers were not even retainable, which finally left me with no more than reporting a range from 4% to 90% regarding the response rate.

5.4.4 *Participants*

One hundred and sixteen German patients and 111 US patients fully completed either the online version or the paper version of the questionnaire. As versions of the questionnaire were distributed not only by paper but also by email and by webpages to patients across both countries, participants of all regions of both countries should be captured.

For the German sample, patients who participated in the study are considered to be representative of the community of cancer patients, as not only patients having internet access were reached but also a proper number without this access ($N = 55$) and hospitalized patients at different sites. For the US sample, I see some restrictions in interpreting later results. Not only have I gathered data mainly from patients having internet access ($N = 100$), but also from patients who are all members of a patient group (see *Summary – Discussion*). Therefore, results may not be generalizable to the general cancer population. The only condition that had to be fulfilled for participating in this study was to be diagnosed with cancer, as then patients were expected to have had to face a decision on a therapy, which ultimately forms the basis for facing a decision on any type of test including pharmacodiagnostic ones.

To begin with the German sample ($N = 116$), patients' ages ranged from 26 to 86 years, with a mean of 54.11 years. The majority of participants were female ($N = 101$), while 15 participants were male. Forty-five patients were high-school educated, 39 had a university or graduate degree, 20 an associated or college degree, while 12 were less than high-school educated. Regarding the type of cancer from the 116 German patients, most patients were diagnosed with mamma carcinoma ($N = 72$), and colon cancer ($N = 17$), while for the rest of the sample only small cancer subgroups were identified (see Table 5.6). The major therapies administered to participants were chemotherapy ($N = 82$), followed by surgery ($N = 81$),

radiation (N = 55), hormonal therapy (N = 33), nonspecified other therapies (N = 17), and targeted therapy (N = 6). The desire for participating in the decision on their treatment was quite high (N = 95), but only 69 of the German participants felt that they finally had the opportunity to do so. Most of the patients looked for further information (N = 97), mainly in books but also in the internet, and a further doctor's opinion was used quite often, while seeking for information within patient groups occurred quite rarely in the German sample. The group was fairly homogenous regarding concerns about the potential side effects of cancer treatments, 89 out of 116 reported having side effects, but rather heterogeneous with respect to the extent of these. While for the majority of patients the concerns about side effects would not make them refuse the treatment (N = 63), for 27 patients it was. Seventy-six felt that they had a supporting and positive relationship with their oncologist; 31 regarded this as being neutral, and only 10 as cold and reserved. Money spent on additional over-the-counter drugs or alternative treatments ranged from €0 to €1,000, while for the majority this was under €100.

For the US sample (N = 111), patients' ages ranged from 24 to 91 years, with a mean of 52.57 years. Ninety were female and 20 male. Again, the majority of patients were university educated (67), 29 were associated with a college degree, 12 had a high-school degree, while only 2 participants were less than high-school educated. Out of all 111 US patients, again the majority of participants were diagnosed with mamma carcinoma (N = 73), followed by ovarian cancer (N = 9), prostate cancer (N = 7), hematological tumors (N = 7), as well as colon cancer (N = 5). At the metastatic stage, 63 patients were diagnosed at the time the study took place. Quite similar to the German study, the majority of the patients had been treated with chemotherapy (N = 61) and surgery (N = 62), while again only a minor number of them (N = 2) were administered targeted therapy. The desire for participating in the decision on their treatment was high, 108 of the US participants wanted to participate, and a great number felt they finally did so (N = 100). Each participant of the US sample had looked for further information, mainly with friends (N = 64), but also a second opinion by another doctor and the internet (N = 51) were used quite often. Again, the group was also fairly homogenous regarding the general concerns about the potential side effects of cancer treatments (N = 90), but rather heterogeneous with respect to the extent of these. For the majority of patients, the concerns about side effects were not strong enough to refuse treatment (N = 62), although for 28 participants it was. Regarding the

Table 5.6: Demography and Intragroup-Related Values of the Patients' Main Study

Questions	Values	Germany (N = 116)	USA (N = 111)
D1: How old are you?	Open field	26–86 years (mean: 54.11; SD: 11.51)	24–91 years (mean: 52.57; SD: 10.87)
D2: What is your gender?	– Female	N = 101	N = 90
	– Male	N = 15	N = 20
D3: What is your highest level of education?	– University, bachelors, graduate degree	N = 39	N = 67
	– Associate degree, community college	N = 20	N = 29
	– High school graduate	N = 45	N = 12
	– Some high school	N = 11	N = 2
	– None	N = 1	N = 0
D4: Which kind of cancer did/do you have?	Open field	N = 72 breast cancer N = 1 prostate cancer N = 17 colon cancer N = 6 hematological cancer N = 2 lung cancer N = 9 liver/bladder cancer N = 6 ovarian cancer 3 missing value	1 missing value N = 73 breast cancer N = 7 prostate cancer N = 5 colon cancer N = 7 hematological cancer N = 2 liver/bladder cancer N = 9 ovarian cancer N = 2 ZNS tumors N = 2 pancreas cancer N = 2 skin tumors 2 missing value
D5: At which stage was the tumor at the time of (last) diagnosis?	– Metastatic	N = 62	N = 63
	– Nonmetastatic	N = 51	N = 48
D6: What type of treatment have you received? (please check all that apply.)	– Chemotherapy	N = 82	N = 61
	– Hormone therapy	N = 33	N = 27
	– Antibody/target therapy	N = 6	N = 2
	– Surgery	N = 81	N = 62
	– Radiation	N = 55	N = 50
	– Other	N = 17	N = 19
D7: When your doctor decided on your treatment, did you wish to participate in the decision?	– Yes, I did. (continue with question D8)	N = 95	N = 108
	– No, I did not. (continue with question D10)	N = 12	N = 0
	– I do not remember if I wanted to participate in the decision. (continue with question D10)	N = 9	N = 3
D8: Did you feel your doctor gave you the opportunity to take part in the decision about your treatment?	– Yes, they did. (continue with question D9)	N = 69	N = 100
	– No, they did not. (continue with question D10)	N = 24	N = 7
	– I do not remember. (continue with question D10)	N = 2	N = 1
D9: To what extent were you eventually	– My doctor gave me a recommendation; the final		

able to participate in the decision?	decision was up to me.	N = 28	N = 63
	– The decision was made jointly with the doctor.	N = 41	N = 37
	– I do not remember.	N = 0	N = 0
D10: Upon receiving your diagnosis, did you try to get more information on your disease and/or your treatment options?	– Yes. (continue with question D11)	N = 97	N = 108
	– No. (continue with question D12)	N = 19	N = 3
D11: Which sources did you turn to? (please check all that apply.)	– Books/journals	N = 61	N = 38
	– Internet/media	N = 55	N = 51
	– Consulting other specialists	N = 51	N = 44
		N = 20	N = 31
		N = 27	N = 64
	– Patient groups		
	– Friends/family		
D12: Before treatment, did you worry about the potential side effects of the therapy?	– Yes, I did worry. (continue with question D13)	N = 89	N = 90
	– No, I was not very concerned. (continue with question D14)	N = 27	N = 16
	– I do not remember whether or not I was concerned.	N = 0	N = 0
	(continue with question D14)		
D13: How strong was your concern about side effects?	– I was so concerned that I refused or almost refused a treatment recommended.	N = 27	N = 28
		N = 37	N = 49
	– I was concerned, but would not have refused a drug recommended by my doctor.	N = 25	N = 13
	– I was concerned to a minor extent, but this did not play any decisive role.		
D14: Which of the following best describes your relationship to your doctor? (please check all that apply.)	– The relationship with my doctor was trusting and positive. They had taken my worries and fears seriously.	N = 76	N = 98
	– The relationship with my doctor was neutral. They had not always taken my worries and fears seriously.	N = 31	N = 9
	– The relationship with my doctor was negative/cool/reserved. They had		

	hardly/never taken my worries and fears seriously.	N = 10	N = 9
D15: How much have you spent in the past for additional over-the- counter drugs? (e.g., alternative treatment) (€/ \$ average per month)	Open field	0–1.000 € (mean: 73.16; SD: 119.45)	0–2.000 \$ (mean: 83.06; SD: 216.97)

Note: Numbers do not always add up to the overall sample size, since not all questions had to be answered depending on the respective previous answer as well as due missing data

relationship to their oncologists, 98 out of 111 felt that they had a supporting and positive relationship with them; 9 regarded this as being neutral, and 9 as cold and reserved. Money spent on additional over-the-counter drugs or alternative treatments ranged, with the US participants, from \$0 to \$2,000, while again it was less than \$100 for the majority.

Between the German and US sample, significant differences were found regarding the gender proportion, education, as well as the therapy options of chemotherapy, surgery, and targeted therapy. Furthermore, the desire and, consequently, extent of involvement felt regarding the treatment decisions taken differed significantly between the groups, with a higher desire on the side of the US sample. The same was found for the search for information with friends and patient groups.

Finally, the number of people who reported having felt their relationship to their doctor as being positive and trustful was significantly higher in the US sample. Table 5.6 summarizes the characteristics of the German and the US sample.

5.4.5 Analysis Considerations

Again, the upfront aim was to compare a noncompensatory fast and frugal model with the compensatory logistic regression model, but, also here, this was finally not possible due the low case-to-cue ratio (3:1) in this study. Therefore, again, the two other compensatory integration models, namely the *Franklin's rule* and the *Dawes' rule*, were chosen as competitors for the noncompensatory fast and frugal model *Matching Heuristic* (Dhami & Ayton, 2001).

The procedures for modeling the patients' test decisions with each of the three models (Franklin's rule, Dawes' rule, and Matching Heuristic) were exactly the same as outlined for the oncologist main study (see Chapter III). Again, as I was not able to determine an outside-

criterion with respect to whether the choice made is correct or not as it was in research of e.g., Gigerenzer and colleagues (1996, 1999) (for more details, see Chapter III: *Analysis Considerations*). Therefore, in analogy to what is commonly reported as cue validity for models such as Franklin’s rule, I determined for each cue its importance⁶³.

Also here, polytomous cues derived from the patients’ pilot study were dichotomized for ease of the analyses, and for each cue all italicized values were coded as “0” and nonitalicized values were coded as “1” (see Table 5.7). The dichotomization was again based on information collected within the pilot study interviews as well as on the weights of each cue level derived from an overall multiple regression carried out on data per country group. As the choice was already of binary nature, there was again no need to simplify it.

The test decision for each patient was modeled on a set of nine cases presented to each participant. Again, this model, with the best prediction in terms of percentage of correctly predicted unknown cases (*k*-1 cross-validating), was chosen as the model for the patients’ test application policy (see Chapter III: *Analysis Consideration*).

Table 5.7: Dichotomized Cue Values of Patients’ Main Study

Cues	Level	Distribution Part A	Distribution Part B
Severity of the side effects of the treatment⁶⁴	(0) Therapy X causes moderate side effects in the majority of patients. 10 out of 100 patients may even have serious side effects.	(0) 5	(0) 4
	(1) Therapy X causes serious side effects in the majority of patients. 5 out of 100 patients may even have life-threatening side effects.	(1) 4	(1) 5
Cost of the test	(0) €/ \$ 2,000	(0) 3	(0) 3
	(1) €/ \$ 300	(1) 3	(1) 3
	(2) Covered by health insurance	(2) 3	(2) 3
Test recommendation by oncologist	(0) Your doctor strongly advises against using the test, because s/he has doubts about the quality of the test.	(0) 3	(0) 3
	(1) Your doctor does not see any reason to use the test but also no		

⁶³ Please be aware of the fact that patients were faced with a binary “yes”/“no” choice, that is, only one object of choice is presented and not two or more. Therefore, each cue’s importance weight is not based on discriminative values of a cue, as it would be the case when at least two objects of choice were presented (see Chapter IV).

⁶⁴ On the webpage, patients were provided with a clickable overview comprehensively explaining what the different side-effect levels mean.

reason not to use it, so the decision is left to you.	(1) 3	(1) 3
(2) Your doctor strongly recommends using the test, because s/he thinks it is of high quality and provides important information for use in your treatment.	(2) 3	(2) 3

5.4.6 Results

Over the course of the next pages, I am going to start first by giving a short overview about main study participants' overall choice behavior before I continue with an outline of the hypothesis-based findings of this study.

Should I say “yes” or “no”?: Patients' overall choice and agreement between them. Overall, in 54.77% of the cases patients decided to have a pharmacodiagnostic test, while within the German sample (N = 116) 56.70% did so and within the US sample (N = 111) 52.70% of the participants. Over the set of nine choices, German patients opted for a test 5.11 times, on average, while their American counterparts did so 5 times. This difference between the two countries was small (Field, 2005) and not significant ($t(225) = .394, p = .684, r = .03$). Figure 5.1 shows the respective results.

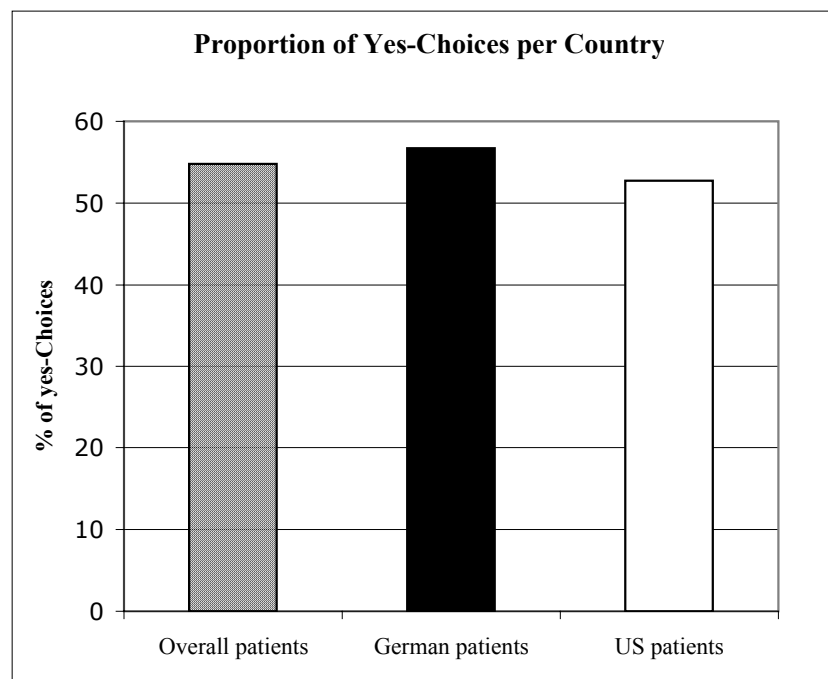


Figure 5.1: Proportion of yes-choices shown for all patients as well as separated for each country.

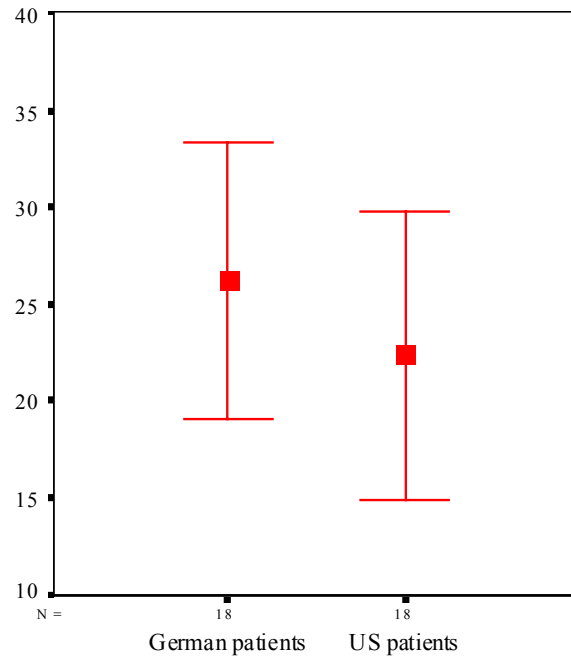


Figure 5.2: Disagreement (mean: 95% CI) over all 18 cases (9 per version) of the German patient sample as well as the US patient sample.

The extent of disagreement was also calculated for the patient study in percentage of their disagreement with the modal choice on each case (for more details, see Chapter III: *Results*). Within each sample, I found different extents of disagreement between patients regarding the decision to be made on each of the nine cases. The disagreement for the German sample ranged from 1.64% to 47.27% (mean: 26.21; *SD*: 14.31), while for the US sample it ranged from 1.79% to 49.09% (mean: 22.33; *SD*: 14.87). This difference in the disagreement of patients between the two countries was small and found to be not significant ($t(34) = .78$, $p = .431$, $r = .14$). Figure 5.2 exhibits the mean and the 95% confidence interval for the data per country.

Happy fitting?: Description of patients' test decision policy. In the previous paragraph of the main study, I hypothesized that a fast and frugal model will provide a better fit to the patients' test decision-making data than a compensatory integration model, such as regression (H4a). Identical to the oncologist main study based on design limitations, I opted for Franklin's rule and Dawes' rule as representatives of a compensatory integration model instead of a

regression model (see also Chapter III: *Analytic Consideration*) and for the Matching Heuristic as the representative of a fast and frugal model.

Given the binary nature of the patients' task, any valid model should be expected to perform better than chance, that is, should predict more than 50% of the decisions. With respect to which model out of the three applied fitted a single participant's choices best, Matching Heuristic as well as Franklin's rule provided the best fit for quite large groups in each of the samples, with Matching Heuristic for 35.3% of the German and 36% of the American patients, and Franklin's rule with 33.6% and 29.7%, respectively. Dawes' rule, however, fitted data best for only 2.6% of the German and 1.8% of the US patients. For the remaining 28.4% of the German and 32.4% of the US participants either two or all of the three models were of an equal overall fit for a participant's choices. Again, it is worth noting that these findings do not necessarily have an implication on the general achievement of a model to fit data well, on average. A model can have an achievement of 100%, that is, fits all choices of a participant correctly, but shares this overall fit with another model. Thus, the previous results give information about how sensitive a model is to fit a single participant's choices on average better than the other two models, that is, to share least best overall fits with other models. Figure 5.3 illustrates these findings. I tried to find out, furthermore, whether differences, in which model the data fitted best, was explainable by any intragroup differences.

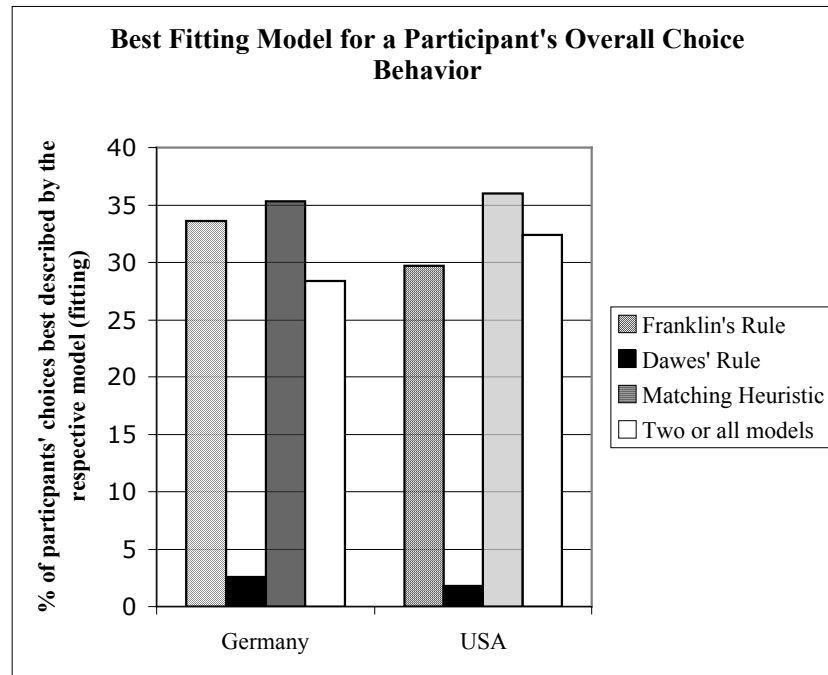


Figure 5.3: *Proportion of the best fitting model for a participant's overall choice behavior—provided by either of Franklin's rule, Dawes' rule, Matching Heuristic or by two or all models for the German and the US patients.*

Again, for both samples, calculating a Chi-square test was not possible with Dawes' rule included, as the respective cells showed expected frequencies below 5.0, which violated one of its major assumptions. When I investigated only the participants best fitted by either Franklin's rule or Matching Heuristic, medium effect sized significant differences were found exclusively for the variable investigating the degree of past concerns about side effects for the German ($\chi^2 (3) = 14.18, p = .003, \phi = .42$) as well as for the US sample ($\chi^2 (3) = 9.14, p = .027, \phi = .35$). Additionally, I found, with respect to the best fitting model, furthermore, for the US sample, a medium sized and significant dissimilarity for this variable, which investigated if any concerns had even existed about side effects ($\chi^2 (2) = 8.52, p = .014, \phi = .34$). Significantly more participants were better fitted by the Matching Heuristic model who were either concerned about side effects and/or within the group of the concerned participants with a higher degree of concern.

Over all participants' choices, the average fit of the Matching Heuristic model for the German sample was 81.90% (range = 44%–100%) and for the US sample 82.68% (range = 33%–100%). For the Dawes' rule, the average fit of the model was 73.95% (range = 44%–100%) for German patients and 74.78% (range = 44%–100 %) for their US counterparts.

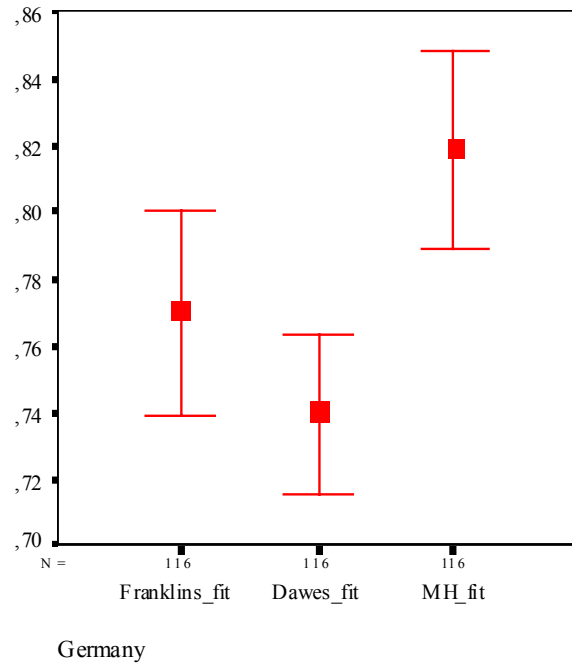


Figure 5.4: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German patients (N = 116).

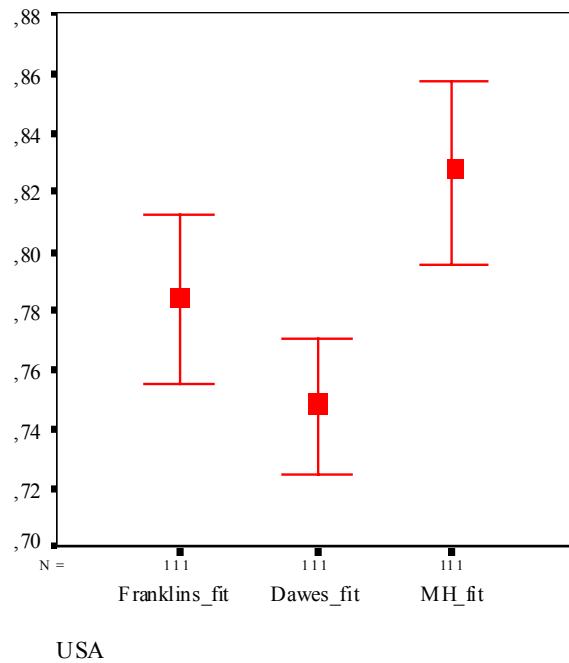


Figure 5.5: Results of average fit (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the US patients (N = 111).

Franklin's rule reached an average fit for the German sample of 77.01% (range = 11%–100%) and 78.33% for the American sample (range: 33%–100%). Figures 5.4 and 5.5 exhibit the respective results per country.

Within the German sample, results showed a significant difference in the degree of fit over individuals between each model in favor of the Matching Heuristic ($F(1.70/195.93) = 12.53, p = .000, \omega^2 = .12$). Mauchly's test, however, indicated that the assumption of sphericity had been violated ($\chi^2(2) = 21.77, p = .000$); therefore, degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\varepsilon = .85$). For the US sample, identical results were found ($F(1.86/204.17) = 15.83, p = .000, \omega^2 = .09$). Also, for this sample, the assumption of sphericity had been violated ($\chi^2(2) = 8.80, p = .000$); for this reason, degrees of freedom were again corrected using the Greenhouse-Geisser estimates of sphericity ($\varepsilon = .93$). With respect to the question of whether any intercultural differences existed, results elicited very small sized and nonsignificant differences when comparing each of the three models' mean fit with its counterpart in the other country (Franklin's rule: $t(225) = -.65, p = .518, r = .04$; Dawes' rule: $t(225) = -.50, p = .619, r = .03$; Matching Heuristic: $t(225) = -.37, p = .715, r = .02$).

Given these findings, I consider Hypothesis 4a as being confirmed for both the German and the US sample, in that the Matching Heuristic model provided not only an equal fit but even a better fit to the patients' test decision-making data than both of the compensatory integration models.

How general are the models?: Generalization performance of the three models. A model's superiority in terms of providing a picture of reality, however, is rather measured by its generalization performance than by its fitting. I therefore again courageously hypothesized that a fast and frugal model would also provide an equal prediction to unknown data (generalization), as would the two compensatory integration models (H4b).

The Matching Heuristic was able to enhance its superiority, compared to the fitting situation, and delivered, for most of the patients of both samples, the best prediction to unknown cases (generalization), that is, counted most hits for a single

participant, compared to the other two models.⁶⁵ Within the German sample, for 43.1% of the participants, their unknown cases were best predicted by this, within the US sample it was even 54.1%. In contrast to its performance in the fitting setting, for the generalization task, Franklin's rule was merely able to predict 14.7% of the German and 12.6% of the American unknown choices better than the other two models. Dawes' rule was found to be least able to predict unknown choices better than the other two models—it did so for 9.5% of the German and 7.2% of the American participants. For the remaining 32.8% of the German sample as well as for 26.1% of the US sample, either two or all of the three models provided an equal performance in an individual overall prediction. Figure 5.6 illustrates these results for each model and country. Regarding the investigation of any intragroup differences, I found a comparable starting situation, as already reported in the “fitting” part. However, here I was not even able to calculate a Chi-square test for at least specific models, since as a minimum for each variable, some cells related to Franklin's rule, and Dawes' rule violated the expected frequency assumption, and, in a number of these cases, even cells related to the Matching Heuristic. The only variable for which it was possible to calculate a Chi-square was the “cancer stage,” which was found not to differentiate between participants whose choice behavior is better by one model, but not by the other.

⁶⁵ Note that low proportions do not necessarily imply that the respective model has a low-prediction performance, on average. Instead, results refer to a model's ability to predict the overall choice for a single participant better than the other two models, and, therefore, does not share an equal performance with another model.

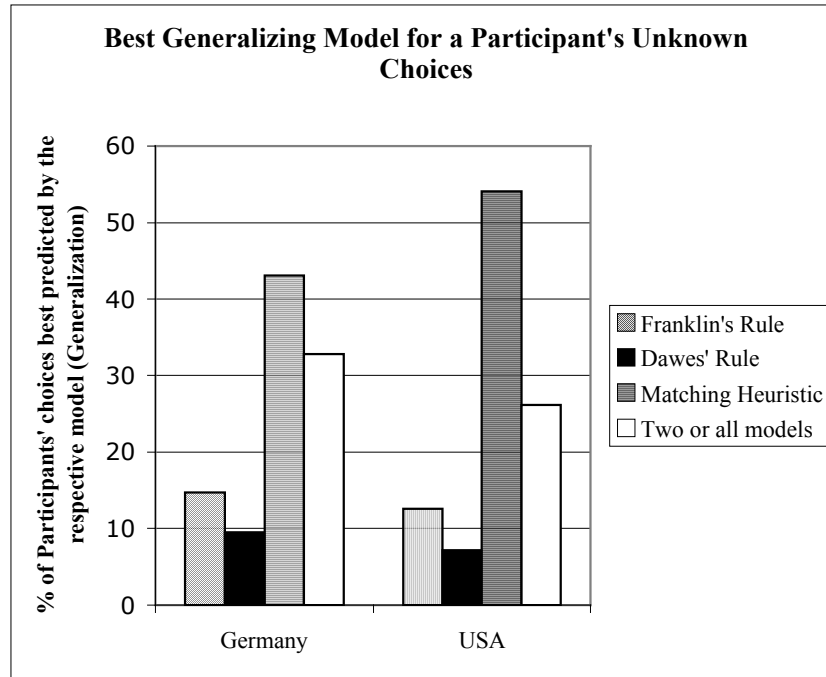


Figure 5.6: Proportion of the best generalizing model for a participant's overall choice behavior—provided by either Franklin's rule, Dawes' rule, Matching Heuristic or by two or all models for the German and the US patients.

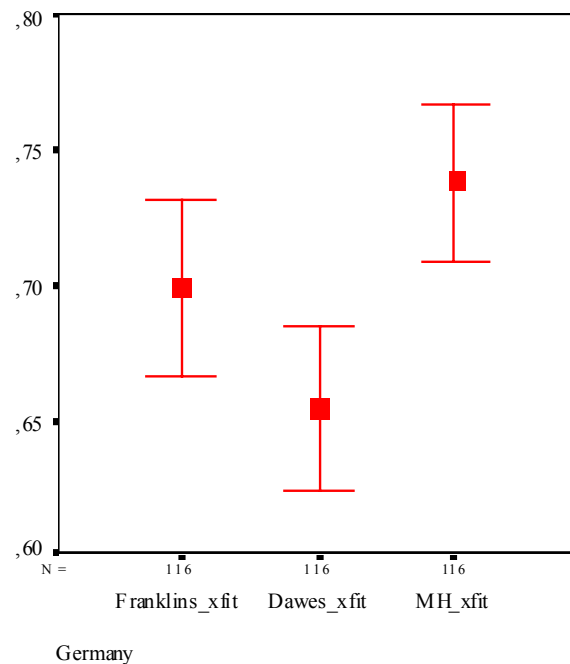


Figure 5.7: Results of the average generalization performance (mean: 95% CI) of Franklin's rule, Dawes' rule, and Matching Heuristic for the German patients (N = 116).

Across all participants, the Matching Heuristic provided the best average of generalization for 73.85% (range = 33%–100%) of the German sample and for 77.98% (range = 22%–100%) of the US sample. With respect to this, Franklin’s rule average prediction was 69.83% (range = 11%–100%) for the German choices and 72.87% (range = 33%–100%) for the US choices.

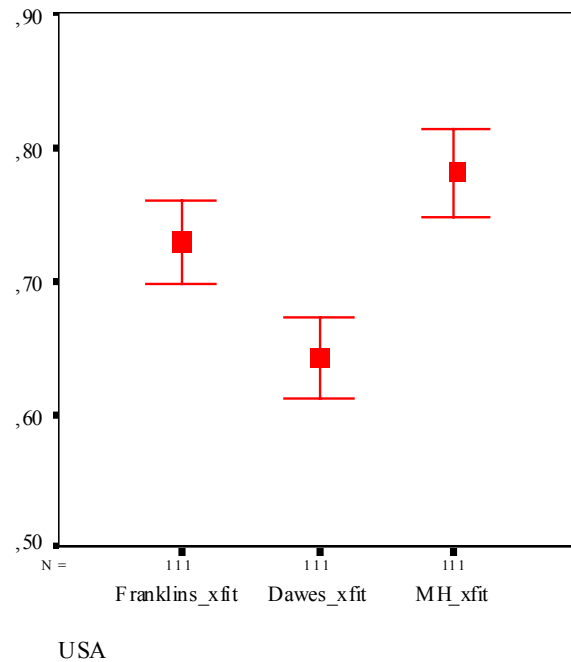


Figure 5.8: Results of the average generalization performance (mean: 95% CI) of Franklin’s rule, Dawes’ rule, and Matching Heuristic for the US patients ($N = 111$).

Again, Dawes’ rule provided the lowest average in terms of predicting unknown cases with 65.42% (range = 33%–100%) for the German sample’s choices and for the US sample with 64.16% (range = 22%–100%). Figures 5.7 and 5.8 give an idea about the mean and the 95% confidence interval for each model’s generalization performance per country. Within the German sample, differences in the average generalization performance of the three models indicated to be small sized and significant between each of the models ($F(2/230) = 11.48$, $p = .000$, $\omega^2 = .19$). For the US sample, identical results were found ($F(1.85/203.09) = 33.55$, $p = .000$, $\omega^2 = .24$). However, for this sample, Mauchly’s test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 9.48$, $p = .009$); for this reason degrees of freedom were again corrected by using the Greenhouse-Geisser estimates of sphericity ($\varepsilon = .92$).

Regarding the interest in any cultural differences, I found that between both countries, differences in the models' average generalization performance were small and nonsignificant for all of them (Matching Heuristic: $t(225) = -1.85, p = .066, r = .12$); Franklin's rule: $t(225) = -1.35, p = .178, r = .08$); Dawes' rule: $t(225) = .58, p = .566, r = .04$).

For the prediction setting, I found the fast and frugal model to even better predict the patients' data than either of the two compensatory models. On the basis of these findings, which are in accordance with the finding of the fitting task, I see Hypothesis H4b as being confirmed, and, therefore, conclude that a fast and frugal model better predicts the choices of patients for a task such as presented than compensatory integrative models.

Taking all or one?: Patients' cue use. Again, I was interested in learning about the cues that participants had used to arrive at their choice. The Matching Heuristic has shown to be best in fitting and generalizing patients' data, therefore, over the course of the next paragraph, I will give an insight into how many and what cues were seen to be important.

I defined cue use already in former chapters (e.g., see Chapter IV) broadly as the number of cues searched for, which includes the cue on which the final decision is based. As I stated earlier (Chapter IV), the Matching Heuristic is not a static model, that is, while a participant might use three cues in one case, they would use only two in another, which is why the number per case may vary.

Within this study, the number of cues used over all nine cases was calculated for each patient. Across all participants, the mean number of cues used ranged from 1–3 (*mean*: 1.23; *SD*: .50). Although identical in range, the German cue usage mean was somewhat higher (*mean*: 1.33; *SD*: .73), while this was slightly under the overall mean (*mean*: 1.16; *SD*: .51) for the US sample. This difference in mean between both countries was small and not significant ($\chi^2(9) = 12.07, p = .209, \phi = .21$). Taking a closer look at the number of cues used, it was found that within the German sample 82.76% of participants' choices were explained by a K=1 model, 6.90% by a K=2 model, and 10.34% by a K=3 model, respectively. For the US sample, a K=1 model was appropriate for 89.19% of participants' choices, while for 8.11% of them a K=2 model predicted data best, and for 2.70% a model using all three cues (K=3).

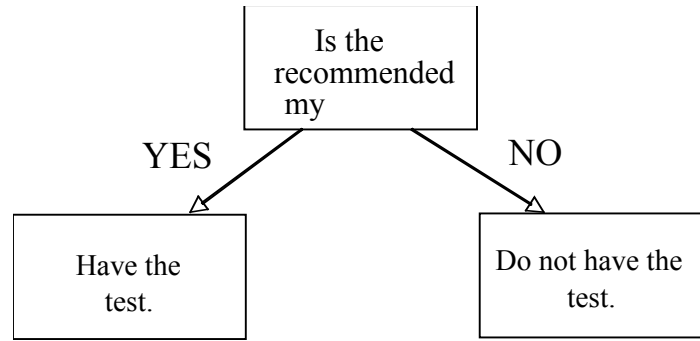


Figure 5.9: Example of a $K=1$ model for a US patient displaying the most frequently used cue within this sample.

Comparing the Matching Heuristic across the participants of each country, it was found, within both samples, that the applied models differed in terms of the cues they used to make their decision. With 64.66%, the information about the cost of the test was most often used by German participants, followed by information about whether the oncologist would recommend doing the test (45.69%) and information about the side effects that the related treatment would cause (22.41%). For the US participants, however, the recommendation of the test given by the oncologist were most frequently used, here by 67.57%, followed by information on costs (35.14%), but least used was the information about the potential side effects of the therapy (13.51%). Figure 5.9 gives an example of a $K=1$ model for an American participant.

Reported representativeness of the questionnaire. With respect to the felt representativeness of the questionnaire, the majority of both German as well as American patients regarded answering the questionnaire easy from start to end, and found the described situation straightforward. In addition, they considered the offered decision criteria in the scenarios also as those criteria that they would use when faced with such a decision. Only a minority of the patients of each sample requested further information (see next paragraph: *Requests for further information*). The time participants needed to work through the scenarios ranged from 2 to 60 minutes (mean: 12.40; SD : 10.95) for the German sample, and for the US participants from 1 to 28 minutes (mean: 7.67; SD : 4.60). Investigating for differences between the two samples with a Chi-square test for the variables, $F1$ to $F3$, only a significant, although small, difference for variable $F3$ was found between the two samples ($\chi^2(1) = 5.98$, $p = .014$, $\phi = .16$). In addition, both samples differed to a medium significant extent from one another when

investigating the time needed to work through the scenarios of both samples ($t(154.21) = 4.26, p = .000; r = .32$). Table 5.8 exhibits all respected values in detail.

Table 5.8: *Values of Reported Representativeness of the Patient Questionnaire*

Question	Germany (N = 116)	USA (N = 111)
F1	(1) N = 75–64.7% (2) N = 36–31.0% (3) N = 5–4.5%	(1) N = 81–73.0% (2) N = 25–22.5% (3) N = 5–4.5%
F2	(1) N = 106–91.4% (2) N = 10–8.6% (3) N = 0	(1) N = 91–82.0% (2) N = 19–17.1% (3) N = 1–0.9%
F3	(1) N = 111–95.7% (2) N = 5–4.3% (3) N = 0	(1) N = 96–86.5% (2) N = 15–13.5% (3) N = 0
F4	N = 5–4.3% See below (requested information)	N = 15–13.5% See below (requested information)
F5	2–60 minutes (mean: 12.40; SD: 10.95)	1–28 minutes (mean: 7.67; SD: 4.60)

Requests for further information. Within the German sample a total of 5 requests, and within the US sample a total of 15 requests, for more information were made by patients when asked if they missed information within the scenarios that they would normally use for decisions (F4). For the German sample, 60% of these 5 patients requested information for a second opinion from another doctor, 20% wanted to have more detailed information of the type and duration of the therapy's side effects, and another 20% requested further information regarding how trustable the results of the test may be.

For the US participants, of those 15 requesting further information, 86% requested information for a second opinion from another doctor, 7% of these patients wanted to know more about the type of side effects of the therapy, while another 7% requested more details regarding the benefit rates of the treatment.

Further findings beyond fitting and predicting: Within the pilot study section, I proposed some further hypotheses beyond fitting and predicting, which I developed from findings within this pilot study. In the following, I will present the results for these hypotheses.

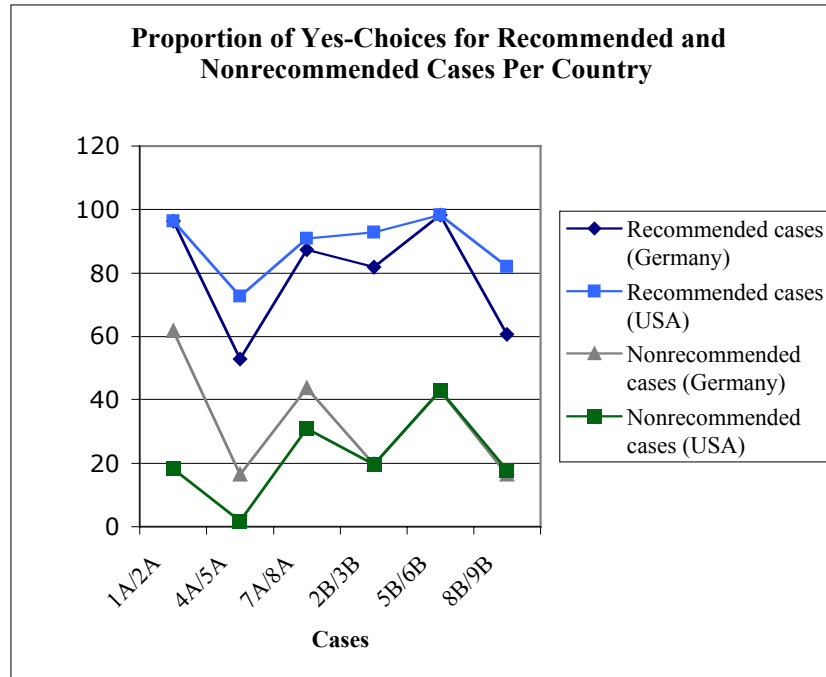


Figure 5.10: Proportion of yes-choices for recommended and nonrecommended cases per country (H1a).

Our very first hypothesis (H1a) was that if a pharmacodiagnostic test is recommended by a patient's oncologist, the patient would significantly more opt for having the test than if it was not recommended. Taking a glance at the cases where the recommendation cue obtained the value for the recommendation by patients' oncologists,⁶⁶ I found that, across these cases, 52.73% to 98.38% (*mean*: 79.56, *SD*: 18.86) of the German patients and 72.73% to 98.21% (*mean*: 88.87, *SD*: 9.69) of the American patients opted for the test when recommended by their oncologist. For those cases that did not receive the respective recommendation cue value,⁶⁷ the patients' decision to have the test anyway ranged in Germany from 16.32% to 61.82% (*mean*: 33.42, *SD*: 18.79) and in the USA from 1.82% to 42.86%, respectively.

When comparing these cases with the respective recommendation value with those cases without the value for the German sample, I saw a small difference, which was not significant ($t(115) = 1.36, p = .087; r = .13$). Within the US sample, however, the difference between the choice for cases with and without the respective recommendation cue was

⁶⁶ These cases were: 1A, 4A, as well as 7A for the A version of the questionnaire, and 2B, 5B, and 8B for the B version.

⁶⁷ These cases were: 2A, 5A, as well as 8A for the A version of the questionnaire, and 3B, 6B, and 9B for the B version.

medium sized and highly significant ($t(110) = 2.86, p = .003; r = .26$). Comparing both countries with each other, I found small and significant differences with respect to the frequency for opting for the test when the cases were recommended ($t(203.55) = 2.49; p = .013, r = .17$) for American patients, who opted more frequently for a test under these circumstances. In cases where the test was not recommended, I saw that American patients apparently also followed this recommendation more by deciding less frequently on having the test anyway, compared to their German counterparts ($t(225) = 4.49, p = .000; r = .29$). Given these results, I confirm Hypothesis 1a for the American sample, but not for the German sample. Figure 5.10 illustrates the proportion of yes-choices for recommended and nonrecommended cases per country.

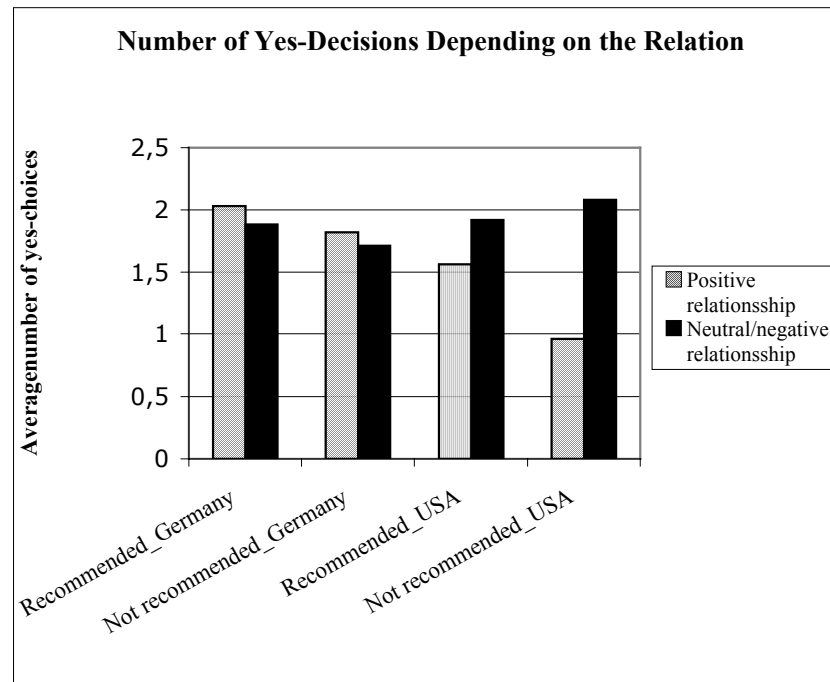


Figure 5.11: Number of yes-choices per country for recommended cases (three per questionnaire version) and nonrecommended cases (three per questionnaire version) depending on the patient's reported relationship to the oncologist (H1b).

I furthermore hypothesized that the intensity of the oncologist-patient relationship could be either a promoter or an obstacle for following the recommendation. In the pilot study, I met patients who reported to have found themselves being less willing to follow their oncologist's recommendation when they missed warmth and trust in the oncologist-patient relationship and experienced the relationship as being reserved. I formulated this idea in Hypothesis 1b. To test this hypothesis, I first aggregated a new variable from the three

relationship variables of the demographic section (see Table 5.5: *Question D14*). Those patients who reported their relationship to their oncologist as being positive received the value “1,” while all those who reported this as having been either neutral or reserved received the value “0.” I then chose all cases having, within its cue profile, the respective recommendation cue value,⁶⁸ and compared the frequency of yes-choices between the two relationship groups. I expected that “0” coded patients should less frequently follow this decision by deciding on having the test. I furthermore chose all cases that included the nonrecommendation cue value,⁶⁹ and investigated if I could find that, in this case, “0” coded patients decide more for having the test than those coded “1.”

For the German sample, I found no difference in the frequency of opting for the test between the two groups, when recommended ($t(114) = .762; p = .224, r = .07$), as in the cases where the test was not recommended ($t(102.84) = .529; p = .229, r = .05$). In accordance with the findings for the situation where the test was recommended, I found no significant difference between the two groups (“0”/“1”) for the US participants either ($t(17.98) = -1.15; p = .133, r = .26$). However, when the test was not recommended by the oncologist, I observed that “0” coded patients decided more frequently to have the test than “1” coded patients. This difference between the two groups was medium sized and significant ($t(108) = -3.32; p = .001, r = .30$). Comparing the “1” coded groups across the countries, I found for both that the situation where the test was recommended ($t(168.77) = 2.62; p = .010, r = .19$) as well as for the cases where the test was not recommended ($t(169) = 4.68; p = .000, r = .34$) small to medium sized differences in the frequency for deciding on the test to be in favor of the German patients. With respect to the frequency for both situations, no differences were found for the “0” coded groups across countries ($t(53) = -.132; p = .895, r = .01$ and $t(53) = -1.18; p = .242, r = .16$). Figure 5.11 illustrates these findings.

In the light of these findings, I only can partly confirm the Hypothesis 1b for the American sample and in cases where the test was not recommended.

Another hypothesis developed on the basis of the findings of the pilot study was the assumption that although patients in both America as well as Germany expressed their willingness to pay for such types of test, in cases where the cost is not covered by the

⁶⁸ These cases were: 1A, 4A, as well as 7A for the A version of the questionnaire, and 2B, 5B, and 8B for the B version.

⁶⁹ These cases were: 2A, 5A, as well as 8A for the A version of the questionnaire, and 3B, 6B, and 9B for the B version.

health insurance, its price could definitely become an eliminating criterion. I expected, especially for German patients, who are still quite unfamiliar with paying for health services privately, that paying for a test would be a larger obstacle than it would be for their American counterparts. It is thus assumed that there is a specific level of test price that will make people decide against the test even if it is recommended, and although the recommendation is expected to remain a paramount criterion for the decision. In order to test this hypothesis, I took all cases where the test was recommended, split these cases regarding their cost values, and compared again the proportion of yes-choices for each value with each other. For the German sample, it was found that each of the three cost levels differed significantly, if rather small sized from one another ($F(2/230) = 60.99$, $p = .000$; $\omega^2 = .16$) with respect to the proportion of yes-choices, while most yes-choices were found when the test price was covered by the health insurance, and least when its cost was €/ \$2.000. Identical results were found for the US patients ($F(2/220) = 21.15$, $p = .000$; $\omega^2 = .06$). Given these results, I assume the Hypothesis 2a as being confirmed. Table 5.9 exhibits the range of proportion, mean, and standard deviation of yes-choices for cases with the respective cost values per country

In the related Hypothesis 2b, I additionally assumed this effect of cost to be stronger for the German sample than for the US sample. However, when comparing the proportion of completed yes-choices for each of the cost levels across the countries, only for the “covered” situation did I find that German patients decided more frequently in favor of having such a test than the US patients ($t(224,63) = 3.47$, $p = .001$; $r = .23$). However, for the actual levels of interest, namely, those where patients were told that they would have to pay for the test, no significant differences for each level was found between the two countries ($t(214.47) = -.44$, $p = .663$; $r = .03$ for the €/ \$300 value; $t(225) = -1.44$, $p = .152$; $r = .10$ for the €/ \$2.000 value). On the basis of these findings, I cannot confirm the Hypothesis 2b, which stated that costs would affect German choices more negatively than the American choices.

Table 5.9: Range of Proportion, Mean, and Standard Deviation of Yes-Choices Exhibited for Each Cost Values and Country

<i>Germany</i>	<i>USA</i>
----------------	------------

<i>Covered</i>	42.62–98.36; mean: 77.19, <i>SD</i> : 21.44	18.18–98.21; mean: 65.98, <i>SD</i> : 32.23
<i>300 €/€</i>	19.67–87.27; mean: 59.23, <i>SD</i> : 25.11	25.00–92.86; mean: 57.75, <i>SD</i> : 32.40
<i>2,000 €/€</i>	16.36–60.66; mean: 34.21, <i>SD</i> : 18.82	1.82–82.14; mean: 40.79, <i>SD</i> : 31.22

Within the last remaining Hypothesis 3, I proposed that particularly concerned patients do not follow an oncologist's recommendation against having the test if the severity of the treatment side effects is high. To assess if this proved true, I chose all cases that had the nonrecommendation cue value "0" and the more severe side effect cue "1." I then compared whether the proportion of yes-choices for having the test differed between patients who reported having been highly concerned about experiencing side effects with those who reported having been moderately concerned (see Table 5.5: *Question D13/1 & Question D13/2*). For both German as well as US patients, I observed a medium sized and highly significant difference between these two groups ($t(62) = 2.58, p = .006; r = .31$ and $t(75) = 2.97, p = .002; r = .32$, respectively). That is, the highly worried patients were found to decide more often for having the nonrecommended test under the condition of a treatment with higher severe side effects than those who were moderately concerned. Figure 5.12 illustrates the respective frequencies depending on the state of concern and per country. Regarding any intercultural differences, no difference was found between neither the moderately worried patients across countries ($t(65.82) = .34, p = .736; r = .04$) nor the highly worried patients ($t(53) = .32, p = .748; r = .04$) with respect to the proportion of opting for a test if not recommended. In the presence of these results, I see the Hypothesis 3 confirmed, in that highly worried patients opt remarkably more often for having a nonrecommended test anyway than less concerned patients.

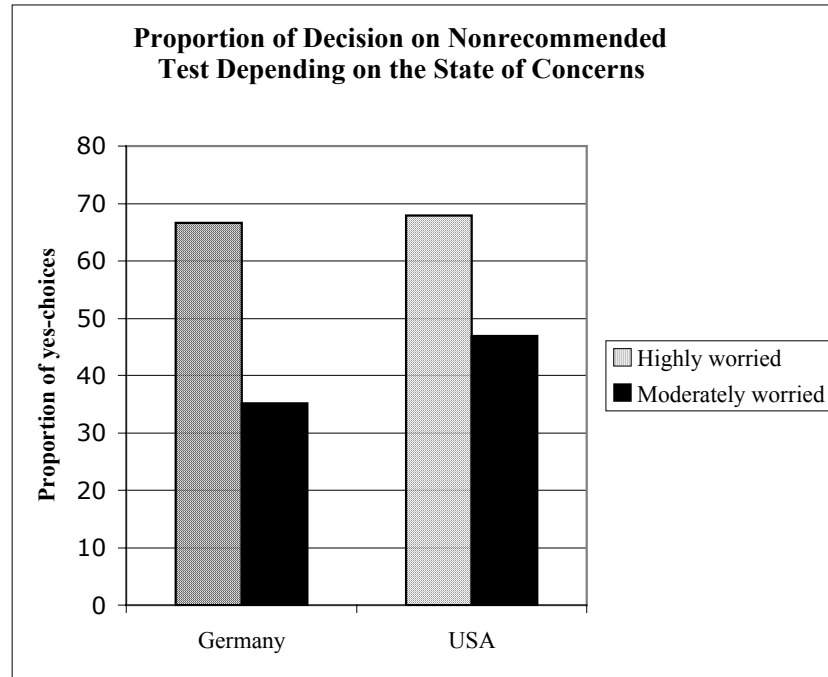


Figure 5.12: *Proportion of yes-choices per country for nonrecommended tests depending on the patient's state of concern regarding the therapy's side effects.*

5.5 Summary & Discussion

The aim of this main study was to learn about patients' cue use and applied decision-making strategies when faced with the opportunity to decide on having a pharmacodiagnostic test before the treatment was administered. Data were derived from nine presented hypothetical cases, which were developed on the basis of findings of a pilot study with cancer patients. For answering the questions of the main study I confronted the data with two compensatory, integrative models, namely Franklin's rule and Dawes' rule, and the noncompensatory Matching Heuristic model. I investigated the relative ability of these three decision-making models to describe and to predict individual patient decisions on a pharmacodiagnostic test made in response to systematically varied case vignettes.

With respect to the general desire of having a pharmacodiagnostic test applied to their treatment decision, for both German and U.S. patients a slight majority of choices in having such a test applied were seen. Patients of both countries also disagreed to a certain extent with regard to the decision to be made on the same case. Furthermore, it was found that German as well as US patients' test decision were better described and predicted by a fast and frugal model called the Matching Heuristic than by either of the two compensatory integrative models, namely Franklin's rule and Dawes' rule. While in addition, US patients opted

significantly more often for doing the test when a recommendation was given by the oncologist than when it was not, such a difference was not noticed for the German sample. In addition, US patients who reported their relationship to their oncologists as being neutral to negative opted considerably more on having a test anyway when their oncologist recommended against doing it than those US patients who reported a positive relationship to their oncologists. For the German sample again such a difference was not found. For both groups, the proportion of opting for the test decreased remarkably with an increase in cost of the test that had to be covered by the patients themselves. Here, German patients were different from US patient in that they opted more for covered tests, even then when oncologists recommended against doing the test. And in addition, there was also at least a tendency for the Germans to decide less frequently than their US counterparts on doing a test, even if recommended, when it had the highest cost value. Finally, it was noticed for those patients who reported to be highly concerned about experiencing side effects that in case side effects were described to be more severe they decided more often on having nonrecommended tests compared to cases where side effects were described as less severe. These patients were found to differ from moderately concerned patients with respect to such decisions.

Although it was shown by several studies that human beings are likely to apply complex decision-making strategies when confronted with decision problems of a high significance and with an outcome not reversible (e.g. Billings & Scherer, 1988; Ford et al., 1989; McAllister et al., 1979), findings of studies done with cancer patients (e.g. Henman et al., 2002; Salkeld et al., 2004) suggested that cancer patients want their oncologists to take the primary role. This was mainly explained by cancer patients' high degree of vulnerability after diagnosis and the feeling to lack expertise to make such significant decisions. This decision behavior is also referred to as example of an expert-opinion-heuristic, following the belief that 'experts can be trusted' (Beisecker, Hlmig, Graham & Moore, 1994; Steginga & Occhipinti, 2002). In that light it might be not too surprising that I found patients' decision making to be best described as well as predicted by a fast and frugal model, namely the Matching Heuristic. 82.8% of the German and 89.2% of the US patients bet on a single reason for making their decision. Nevertheless, my findings were only partly in accordance with the suggestion of such studies. Patients also disagreed to a certain extent with regard to the choice to be made on the same case. This disagreement was based on various reasons such as costs to be covered by the patient himself, concerns about experiencing side effects as well as the recognized

relationship to the oncologists. For the German patients, for instance, it was not the oncologist's recommendation, which most frequently guided their choice. It was the cost of the test, instead. One reason for that might be that it was for a long time commonplace in Germany not to pay for any health services. However, also the German health system has been challenged more and more recently, which has been leading to the situation that now an increasing number of health related services have to be covered by the patients themselves – a fact to which patients only recently have been forced to become adapted to. That the cost cue was found to have such an impact on the choices of the German sample could be an expression of a probably still more negative attitude and reluctance of a majority of patients in Germany to pay for any health services.

In contrast, the US patients, who are used to paying at least for parts of their health services themselves, based their decision mainly on the recommendation given by the oncologist. 65.6% of the 89.2% $k=1$ models were based on that cue. Their reliance in the opinion of the oncologist is not guaranteed, however. The results of this study point towards a decreased likelihood of US patients in following an oncologist's recommendation in case the relationship is not recognized as positive by them. When US patients felt confident that the doctor cared for, understood and respected them, they seemed to be pleased with accepting the doctor's recommendation without engaging in their own decision-making, a finding that is in line with others (Henman et al., 2002; Salkeld et al., 2004). However, in case this was not given, US patients appeared to be quite ready to make up their own mind and did not bother much with what the oncologist deemed the best, a result I could not prove true for the German patients.

Nevertheless, one needs to be cautious about generalizing from these results that German cancer patients do not care about oncologists' recommendations and that the only thing that matters is whether they would have to pay for health services or not. This is surely not the case. Beside the ongoing changes in the health system, which might partly explain the recognized importance of the cost cue, it was additionally noticed that one point in which the German and the US patient groups differed from each other was the degree of education. While the majority of US patients had a university degree, most of the German patients only had a High school degree. The effect of this can be assumed to be at least twofold. First, US patients of this study might have had a higher annual income and therefore, cost issues were not such a significant matter to them as for German participants. Although I did not explicitly

ask patients to reveal the annual income for acceptance reasons of this study, it is likely that better educated people have jobs providing a higher salary. Second, higher educated patients might have a stronger belief and trust in their own judicial abilities with respect to health related issues. This might make them more self-confident in coming up with their own decisions in case an oncologist cannot serve the high need for emotional support and trust. By taking a look at Figure 5.12, it can be seen also for German patients that an oncologist's recommendation makes a difference regarding their choice behavior, even if only a small one.

That is, an oncologist's opinion obviously counts, although not under each condition. An oncologist who is not able to deliver information in a proper way to his/her patient and/or is less able to recognize and address patients' need for trust and information or his/her anxiety regarding side effects may risk that patients made up their own mind and want to see their treatment decision based on a test which has not sufficiently proven to serve scientific demands. This might sound far-fetched right now, as still most of the diagnostic tools including tests are only accessible through a doctor, but within recent years, the pharmaceutical industry has noticed the financial potential in selling 'certainty' over the counter or in pharmacies by offering more and more tests. That is, an oncologist who wants his/her patients in line with his/her opinion has not only to ensure good treatment decisions, but also sufficient emotional support. One might question, however, whether it is smart at all of a patient to transfer a potentially life-altering decision completely to the oncologist.

Although I do not doubt that it can be smart to base a decision on a single reason, it is obvious that the reason has to be a good one in order to finally make indeed a smart decision. Most patients are not specifically educated in medicine and therefore by default less likely than an oncologist to make a proper choice with respect to the most beneficial treatment or an evaluation of when to use a test. It therefore would be rather hazardous to encourage cancer patients to draw their own conclusions when faced with a medical decision and to not care much about an oncologist's opinion. However, in line with the idea of shared decision-making, a smart patient should expect the oncologist to properly inform her/him in comprehensible words about benefits and harms of the considered course of action. Right now, cancer patients commonly tend to look at their oncologists to save their life and frequently defer disease-related decisions, as pilot study oncologists and other reported study findings suggested. This surely is to a certain extent a function of the patients' feeling to lack the ability to understand all the uncertainty of benefits and harms inherent in such kinds of

decision. Clearly, an incorporation of a test does not make things easier. However, I have the strong feeling that this is a two-way street. Although not tested on a larger scale with oncologists, their pilot study results were comparable to results from other clinicians with respect to a proper understanding and interpretation of probabilistic information (e.g., Eddy, 1982; Gigerenzer et al., 1998). If I were assuming that oncologists do not sufficiently understand this construct, however, then it would have to be concluded that they could hardly explain probabilistic information in any comprehensible way to their patients. This would certainly lead to the situation that even if such an oncologist tried to explain it, patients would appear to be puzzled by the probably incomprehensibly explained facts. Thus, it is rather unclear if patients are not sophisticated enough to understand probabilistic facts or clinicians are not prepared enough to explain it in an easy manner. This situation might lead at the end to what I found within the patients' pilot study where no patient was even told about the existence of false test result, which made me finally not include such information within the presented cases. It is like a circle – if there were informed and numerate physicians, which were able to not only to understand probabilistic information, but also to explain such in a comprehensible way, e.g. by using a natural frequencies tree (see Chapter III: *Summary & Discussion*) to their patients, we may find less patients who just want to defer the decision as it appeared too complex to them to get involved in the decision. This, however, is something I would indeed regard as smart: patients who would like to get involved in their medical decisions and physicians who encourage them by offering sufficient information to do so.

In the next Chapter (VI) methodological concerns will be discussed in detail.

6 What Have We Learned (So Far)?: Final Discussion and Conclusion Remarks

6.1 *Model Concerns or the Data's New Clothes*

With this thesis, I have proposed doubts that, in the field of cancer care, the medical decisions of clinicians, such as oncologists and pathologists as well as of cancer patients, are made in line with ideas of unbounded rationality, which assume that people weight cues optimally in a compensatory manner. I questioned the appropriateness of models of unbounded rationality for several reasons, such as they require large attention, memory, and processing abilities. Furthermore, these models ignore the fact that decision strategies are adapted to the demands of the task. However, I outlined that the human mind is characterized by limited cognitive processing capacity, and, in addition, that medical decisions are characterized by conditions, such as task complexity, which have shown to foster the use of more simple noncompensatory decision strategies (for more details, see Chapter II). For this reason, I claimed that, even in the field of cancer care, where usually quite significant decisions have to be made, clinicians as well as patients behave boundedly rational and bet on one single good reason when making their decisions. In order to test whether my assertion would hold, I decided to confront the data of all three studies, that is, oncologists', pathologists', as well as cancer patients' data, with a compensatory integrative model, such as a regression model and a fast and frugal model. Although the regression model would have been a perfect representative of compensatory integrative models within this competition, I noticed, however, some serious concerns connected with the application of the regression approach due to the design of all three studies.

One major concern was that, in applying a regression model to data, one needs to have an adequate case-to-cue ratio in order to arrive at meaningful estimates and conclusions, respectively. Linear models, such as regression models, require a large case-to-cue ratio. For instance, in the case of the oncologist main study, by applying four predictors I would have had to present a single participant with at least 82 cases, otherwise by presenting too few cases⁷⁰ relative to the number of predictor variables a number of problems would have occurred. Under such circumstances, a regression model is assumed to produce, for example,

⁷⁰ A rule of thumb is for calculating the amount of cases: $N \geq 50 + 8 \times \text{number of predictors}$.

large parameter estimates and standard errors as well as predict the depended variable perfectly as an artefact (Tabachnick & Fidell, 1996). Apart from the typically nonexistent willingness of participants to complete such a high number of cases needed for arriving at valid estimates, I also had serious concerns that by torturing participants through a task of such complexity, I would appear to be constructing my own evidence for noncompensatory decision-making strategies. Results from several studies indicate that increasing task complexity (increasing number of alternatives, dimensions, or both alternatives and dimensions) increases the likelihood of subjects to apply more noncompensatory decision strategies (e.g., Biggs et al., 1985; Billing & Marcus, 1983; Johnson & Meyer, 1984; Klayman, 1985; Olshavsky, 1979; Onken et al., 1985; Payne, 1976; Payne & Braunstein, 1978; Staelin, & Payne, 1976). For instance, findings of studies such as those of Dhami and Harries (2001), where general practitioners were found to be equally well-described by a fast and frugal model as they were by a regression model, are vulnerable and open to criticism, when considering that these conclusions are based on the data of 36 participants only, but who were confronted with 130 cases, each made up by 12 cues.

Policies captured by “paper cases” are sensitive to the task presented, and I felt unprepared to sacrifice the reliability of the data in return for the accuracy of the applied statistical model. However, this situation led to the question of, where to get an equivalent for regression from, which would adequately represent the compensatory integrative side of the competition, but not come with the same problems. I found two such variants of weighted linear models, although nonstatistical, in Franklin’s rule and Dawes’ rule, both of which I finally applied within each of the three studies. Franklin’s rule is comparable to a regression model in that it involves the compensatory combination of multiple differentially weighted cues, and is limited in its inflexible cue use. Dawes’ rule is comparable to a regression model in that it is also limited to inflexible cue use and also involves a compensatory combination of all cues, although it employs unit weights and, therefore, proceeds in a less sophisticated manner than regression and Franklin’s rule would. However, both models, Franklin’s rule and Dawes’ rule, are different from a regression model in that they do not compute optimal weights in the least squares sense, nor do they take into account the interdependencies among cues. Since neither of these models, for example, estimate interdependencies among cues such as regression, they do not require such a high case-to-cue ratio in order to arrive at valid results. Several studies have demonstrated that both models are first-rate approximations to

regression models in terms of descriptive and predictive validity (Cattin, 1978; Dawes & Corrigan, 1974; Dorans & Drasgow, 1978; Einhorn & Hogarth, 1975; Gigerenzer & Goldstein, 1999, 1996; Schmidt, 1971; Wainer, 1976; Dhami & Ayton, 2001). For this reason, I consider both models as proper equivalents to a regression model, and, therefore, do not regard the results derived as limited.

6.2 *Implications for Research in Medical Decision Making*

For decades, compensatory integrative models, such as the regression model, have been used by the majority of researchers in applied domains, such as medicine (see Wigton, 1996), in order to capture judgment policies. These models depict human beings behaving in accordance with the ideas of unbounded rationality, in that such approaches assume people to weight all cues optimally in a compensatory way and to use the same cues in the same manner to make decisions on different cases. Researchers have questioned whether such compensatory linear models are capable of providing a psychologically valid description of human judgment, however (Einhorn, 1971; Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Gigerenzer & Goldstein, 1996, 1999; Czerlinski et al., 1999). Despite the criticism that compensatory integrative models are not psychologically plausible due to constraints set by the environment and the human mind itself, the majority of judgment analysis research still uses regression models. Although these linear models usually arrive at quite good estimates in fitting human beings' decisions, it should be stressed that linear models are hardly more than what Hoffman (1960) called a paramorphic representation of human judgment behavior, which might not necessarily be a valid description of the actual reasoning process. A number of explanations for the persistent use of regression models have been highlighted. Reasons offered included arguments, such as an expected higher predictive validity and a more adequate fit by employing regression models (Hoffman, 1960; Stewart, 1988; Brehmer & Brehmer, 1988; Dhami, 2003) as well as that the tools for calculating these models are easily available (Stewart, 1988). Another explanation for the popularity of regression models in policy-capturing research related to the rather low performance of alternative models, for example, conjunctive and disjunctive models (Einhorn, 1970), which were available in the past.

Within recent years, however, several alternative noncompensatory simple models have been proposed (Gigerenzer & Goldstein, 1996, 1999; Dhimi & Ayton, 2001), which have shown to provide an excellent fit and prediction to decision data, and are able to even outperform compensatory model predictions (e.g., Czerlinski et al., 1999, Gigerenzer, & Goldstein, 1999; Dhimi & Ayton, 2001; Dhimi, 2003). For two of the studies conducted here, namely, the study of pathologists and of patients, I found results that foster these findings. While for the pathologists, compensatory and noncompensatory models arrived at equally high predictions, for the patients, the noncompensatory model even outperformed the two compensatory models in its ability to fit and predict decision data. Although in the study of the oncologists the noncompensatory model was not able to outperform both of the compensatory models, it was nevertheless able to outperform one of them, and, in addition, was found to predict still a proper amount of decisions better than either of the compensatory models.

There might have been conditions that may have enabled the two applied noncompensatory models, namely Take The Best heuristic and Matching Heuristic, to prevail in at least two of the studies. However, none of the studies were limited to the meanwhile well-known constraints, such as time pressure or task complexity. In contrast, in the oncologist study, I felt it was not even possible to present the task in its real complexity due to design constraints (see Chapter III: *Main Study–Design*), which might have rather fostered the observed higher degree of application of compensatory decision strategies, compared to noncompensatory ones (see Chapter III: *Results; Summary & Discussion*). Although there were also no constraints on time, it was noticed, nevertheless, that some of the participating oncologists as well as pathologists reported quite short periods of time required to work through the respective questionnaires. While I was not able to test for the oncologists whether such participants' behavior was better explained by a noncompensatory or a compensatory model due to intragroup sample sizes being too small, for pathologists I found no meaningful differences with respect to this issue. However, even if this had supported the application of noncompensatory decision-making strategies, it would depict the clinicians' reality, as time pressure is a common condition that clinicians face within their daily work.

It was additionally considered if the low number of requests for further information, observed within the studies, might have been a consequence of the relatively short amount of

time that clinicians invested in going through the questionnaire. In order to request further information, participants had to make the effort and fill in an open field (see Table 3.5: *Question F4*). Although it is hard to argue for, or against, this concern when setting the findings in relation to those in Question F3, where participants were asked whether they missed information for their choice and which was an easily clickable one, it was found that most of the participants who clicked to have missed something also filled in their request in the next open field question. This might support the finding that participants actually did not have more requests, which should not imply that there had not been information that they should have requested.

With respect to the fairly good performance of the simple noncompensatory models applied here, it has to be highlighted again that these models achieved this performance by usually applying less than half of the presented pieces of information within the respective cases, while both of the compensatory models made use of all the information offered. In addition, the noncompensatory models proposed a flexible strategy, in that they searched only for the necessary cues, which might have differed from case to case. The compensatory models proposed an inflexible strategy, as all cues were used in the same manner. In the light of demands set by the environment, such as time pressure and the limited cognitive processing capacity of human beings, such compensatory models are psychologically implausible. This is clearly a point worth stressing. When choosing a model for capturing humans' decision policies, one must consider how plausible this model will be with respect to truly depicting people's decision policies under the conditions of a restraining environment and human processing capacity. The development of a proper psychological theory of human decision making is possible only if psychologists test the relative predictive validity of such a model that is cognitively plausible. This has been neglected in the past. By being aware of the possible implications that our research of human judgment can have for policymakers, authorities, and the focused professionals themselves, we have to take responsibility for involving and testing cognitively more plausible models under different conditions. This hopefully enables us to arrive at a more valid insight of how people make decisions, and, therefore, help us to provide respective interest groups with more valid advice regarding the application of such results.

6.3 The Issue With Hypothetical “Paper” Cases and Some Other Methodological Concerns

In studies such as presented here, one concern certainly is that the results are based on the decisions made by individuals on a set of hypothetical “paper cases.” The external validity of such cases has been questioned by several researchers (e.g., Gorman, Clover, & Doherty, 1978; Ebbesen & Koneci, 1980; Jones, Gerrity, & Earp, 1990). Nevertheless, after reviewing three decades of policy capturing studies, Brehmer and Brehmer (1988) concluded that the use of a “paper” representation of real cases for capturing people’s decisions did not lead to any important distortions in the policies obtained. This was particularly true in the medical domain, where they argued that judgment policies, captured on the basis of “paper patients,” did provide an externally valid representation of how physicians would judge real patients. Cooksey (1996) arrived at comparable conclusions after considering the internal and external validity of studies using hypothetical cases. Both of these reviews included research using “paper cases” that involved hypothetical cases with orthogonal cues, such as those in the present three studies. Other studies, conducted in the field of medical judgment analysis, support these statements with findings that showed high correlations between decisions made in actual cases and in “paper” ones (Kirwan et al., 1983; Chassin et al., 1987). Thus, I do not consider the present study to be especially limited due to the fact that I investigated decisions made on hypothetical cases consisting of an orthogonal cue set.

All three studies were mainly, and, in the case of the oncologists, exclusively, conducted as online investigations. The usual major concern with respect to such types of investigation is that only policies of specific subgroups of people are captured, and, therefore, the study results lack representativeness. While I regarded having internet access for today’s clinicians, such as oncologists and pathologists, as commonplace, I am quite confident that the representativeness of this study population is not limited due to the online distribution of my research. However, I also shared this concern with respect to the patient samples. While in Germany, I was able to distribute a certain amount of paper versions of the questionnaire to those patients who had no internet access, this was only possible for a very small number in the USA, and, therefore, I mainly used the opportunity of online distribution. This circumstance surely contributes, to a high extent, to the noticed and already highlighted major difference between the two samples (see Chapter V) regarding the education level, with US patients having a higher level.

Another issue that has to be highlighted, relative to the question of the representativeness of my samples, certainly is the difference in recruitment procedures between the German and US sample, in general. While I had the opportunity to receive support from several organizations within Germany in approaching oncologists and pathologists who were from university clinics, hospitals, or private practices, I was not able to receive this same opportunity within the USA. I therefore had to opt for a type of makeshift solution, in that I collected oncologists' addresses from the scientific meeting member directory of ASCO, and, in the case of pathologists, searched university clinics as well as hospital webpages to obtain addresses. In the case of the US oncologists group, I therefore contacted only such oncologists who were more likely to attend a meeting, which, additionally, is regarded as a type of educational training due to its exclusive presentations and workshops. In the case of the US pathologists group, mainly those webpages related to university clinics provided access to their pathology department, which explains the high number of such pathologists coming from the university side within the US sample and the smaller number of those coming from hospitals and private practices. Apart from the already outlined differences in the ability to distribute paper versions of the questionnaire to German and US patients, with respect to the distribution of the online version of the questionnaire I regard the procedure between the German and the US patient sample to be quite comparable. While within the USA I received support from patient groups in distributing the information of the online questionnaire to patients, within Germany several cancer organizations provided identical information for patients on their webpage. With respect to the differences in the respective samples, I checked differences in findings across countries carefully for having possible roots in the differences of recruitment, and, in case this was true, I highlighted this within the respective group-related discussion.

Finally, due to the limited willingness of participants to complete more than ten cases, I opted for splitting the questionnaire into two parts in order to be able to manipulate still a sufficient number of cues while retaining an orthogonal design (see Chapter III, IV, and V: *Main Study – Design*). Each of the cue values was equally distributed among the set of all developed cases as well as among each version of the questionnaire. In addition, on each of the webpages of the respective sample, the questionnaire forms A and B were programmed in such a manner that they were presented randomly to participants, for example, when one participant had worked through version A, the next participant who clicked on the online study

was shown version B, and so forth. That is, a participant could not choose which version to work through as this was fixed by the electronic system. For the paper versions, I took care that the same number of both versions was distributed, and, again, ensured that participants were not able to choose themselves which version they worked through, so that each participant had an equally high likelihood of being presented either version A or version B. For these reasons, I did not feel that it was justified or even plausible to make separated analyses for each of the questionnaire versions for each sample and each country. Given the conditions that cue values was equally distributed among the cases and questionnaire versions, and in addition participants had the same high likelihood of being assigned to either of the questionnaire versions, this procedure has also been suggested within the respective theoretical literature (Louviere, Hensher, & Swait, 2000).

6.3.1 *Some Words on the Matching Heuristic*

While for the patient study it was found that the Matching Heuristic performed equally well in fitting as well as in predicting the data, this was not found for the oncologist study. Here, the Matching Heuristic performed well in the fitting situation, while it did not in the generalization setting. This finding attracted my attention, however.

One characteristic of heuristics is that they are not too specific, which makes these models quite robust. Robustness is the ability to generalize well to new environments, that is, those whose structure is not known in advance. An important distinction here is between data fitting and prediction. If a model shows only good performance in the fitting situation, but not in the generalization situation, then this might indicate overfitting. It is hard to argue, however, for the current study, that the Matching Heuristic overfitted the data, as the model under both conditions showed to have applied not even half of the offered information. Trying to find an answer regarding this observation within other research that had applied this model, it became obvious that, in three studies, only the fitting situation was reported (Dhamsi & Harris, 2001; Kee et al., 2003; Smith & Gilhooly, 2003), and, in the initial study where the Matching Heuristic was introduced (Dhamsi & Ayton, 2001), no significance tests or any other statistical measure e.g., effect size for the comparison with the other models were reported.

This gives rise to several questions with respect to what structure of environment the Matching Heuristic actually exploits, and what knowledge about the predictive value of probabilistic cues it requires. For instance, its performance might be vulnerable or even supported by compounded cues (Garcia-Retamero & Hoffrage, 2006). It is felt that this

question needs to be addressed in the future. Therefore, a simulation study is planned as a follow-up of this thesis.

6.4 *Implications for the Health System*

Results of the oncologist study as well as of the pathologist study highlight, without doubt, the importance of guidelines and experts to the respective groups, while cancer patients were influenced to a rather high extent by the recommendation of their oncologist. That is, everybody relies on somebody else—and to some extent this might even be smart, as nobody is an expert in every field that this world offers. But although the research outlined here did not bear on the accuracy of decisions, there were some points that gave cause for concern.

Oncologists, who chose not to rely on guidelines under different conditions, and pathologists, who appeared not to care much about a higher (analytic) specificity, plus the findings of the pilot studies that showed serious difficulties in understanding test related issues, such as sensitivity and specificity, have given rise to questions about who exactly is in charge of ensuring that only reasonable tests were applied.

It was pointed out that authorities in charge of setting up guidelines are expected to hold at least the necessary knowledge about issues, such as sensitivity, specificity, and predictive values. However, there were several issues stressed with respect to guidelines (see Chapter IV). One of the issues highlighted was that, within Germany, such guidelines are mainly not established yet for pathologists and still under development for oncologists. That is, by now, these two German groups do not have much to rely on. In addition, even in case such guidelines are present, guidelines might not always be such scientifically objective and unbiased tools for guiding clinicians to the best standards of care as would be desirable. Named examples exist (see Chapter IV) which made obvious that guidelines also represent a compilation of certain interests, and it remains rather unclear to clinicians whose “politics” is sold within the respective one.

Given the significant consequences a pharmacodiagnostic test’s result can entail on a cancer patient’s life, I strongly argued to consider serious interventions before several pharmacodiagnostic tests appear on the market. It was highlighted that it has to become the responsibility of each proper medical educational program to ensure that clinicians will be educated in a more comprehensible way in the interpretation and understanding of probabilistic information than has been the case so far with counterintuitive Bayes’ theorem.

For this reason, I recommended the introduction of a simple reproducible natural frequency tree (Gigerenzer & Hoffrage, 1995) to clinicians for resolving the task of understanding and calculating a test's probability, as this approach has been shown to lead to a notable improvement in this respect (see Chapter III). To the extent that clinicians were able to understand and interpret a test's probabilistic information, I assumed, in addition, an increase in clinicians' readiness to involve patients within such decisions, as this understanding would also enable clinicians to explain it properly to their patients. Based on this situation, patients were better informed and better able to make an informed decision.

Nevertheless, even in case clinicians were able to understand and work with probabilistic information properly, decision aids, such as guidelines, could still be regarded as a helpful tool under the given circumstances of a clinician's working day, such as heavy workload and time pressure. Although I would like to argue how important it is to leave "politics" and "interest" amongst others than solely patient's health out of guidelines, I am aware of how quixotic this desire is. Yet, I emphasized that other issues, such as the complexity of guidelines (e.g., Ustun & Sartorius, 1995; Smith et al., 2003) as well as their limited ability to provide the latest developments within medicine (Chapter III: *Pilot Study–Results*), were found to decrease the acceptance of such aid tools. Findings that highlighted, also in the field of medicine, that less is often more (e.g., Forster et al., 2002) should definitely encourage authorities responsible for setting up guidelines to develop these in a manner consistent with models that appear to be used naturally by human beings. It might, for instance, increase the acceptance and usefulness of guidelines when outlining them by using a simple decision tree. Moreover, who knows, by not aiming at compiling long arrays of all available data within guidelines, but concentrating on the most important pieces of information, it might be even possible to keep guidelines updated in a fast *and* frugal way.

The present study differed from other studies conducted in the field of medical decision making in that I investigated decisions on pharmacodiagnostic tests in the field of oncology, where people are faced with potentially life-altering decisions. In addition, this research differed from others done as I tested the decision-making strategies of oncologists, pathologists, and patients with three alternative models of decision making.

References

- Abelson, R.P., & Levi, A. (1985). Decision Making and decision theory. In: G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology – theory and method*. New York: Lawrence Erlbaum Associates, Inc.
- American Cancer Society (2006). *Cancer Facts & Figures 2006*. Received from: <http://www.cancer.org/downloads/STT/CAFF2006PWSecured.pdf> on October, 30th, 2006.
- Arkes, H.R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, 39, 133-144.
- Beisecker, A.E., Helmig, L., Graham, D., & Moore, W.P. (1994). Attitudes of oncologists, oncology nurses, and patients from a women's clinic regarding medical decision making for older and younger breast cancer patients. *Gerontologist*, 34, 504-512.
- Berwick, D.M. (1989). Health services research and quality of care: Assignments for the 1990s. *Medical Care*, 27, 763-771.
- Biggs, S.F., Bedard, J.C., Gaber, B.G., & Linsmeier, T.J. (1985). The effects of task size and similarity on the decision behavior of bank loan officers. *Management Science*, 31, 970-987.
- Billings, R.S., & Scherer, L.M. (1988). The effects of response mode and importance on decision making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, 41, 1–19.
- Billings, R.S., & Marcus, S.A. (1983). Measures of compensatory and non-compensatory models of decision behavior: Process tracing versus policy capturing. *Organizational Behavior and Human Performance*, 31, 331-352.
- Brehmer, B. (1980). In one word: Not from Experience. *Acta Psychologica*, 45, 223-241.
- Brehmer, A., & Brehmer, B. (1988). What has been learned about human judgment from thirty years policy capturing? In B. Brehmer, & C.R.B. Joyce (Eds.), *Human judgment: The SJT view*. Amsterdam (North Holland): Elsevier.
- Brock, D.W., & Wartman, S.A. (1990). When competent patients make irrational choices. *New England Journal of Medicine*, 322, 1595-1599.
- Brody, D.S. (1980). The patient's role in clinical decision-making. *Annals of Internal Medicine*, 93(5), 718-722.
- Butow, P., & Tattersall, M. (2005). Shared decision making in cancer care. *Clinical Psychologist*, 9(2), 54-58.
- Butow, P., Devine, R., Boyer, M., Pendlebury, S., Jackson, M., & Tattersall, M. (2004). A cancer consultation preparation package: Changing patients but not physicians is not enough. *Journal of Clinical Oncology*, 22, 4401-4409.
- Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999–1000.

- Cassileth, B.R., Zupkis, R.V., Sutton-Smith, K., & March, V. (1980). Information and participation preferences among cancer patients. *Annals of Internal Medicine*, 92, 832-836.
- Cattin, P. (1978). A predictive-validity-based procedure for choosing between regression and equal weights. *Organizational Behavior and Human Performance*, 22, 93-102.
- Chapman, G. B., & Elstein, A. (2000). Cognitive Processes and Biases in Medical Decision Making. In: G.B. Chapman & F.A. Sonnenberg (Eds.), *Decision making in health care – Theory, psychology, and application*. Cambridge: Cambridge University Press.
- Charles, C., Gafni, A., & Whelan, T. (1997). Shared decision-making in the medical encounter: What does it mean? - Or, it takes at least two to tango. *Social Science & Medicine*, 44, 681-692.
- Charles, C., Redko, C., Whelan, T., Gafni, A., & Reyno, L. (1998). Doing nothing is no choice: Lay constructions of treatment decision-making among women with early-stage breast cancer. *Sociology of Health and Illness*, 20, 71-95.
- Charles, C., Gafni, A., & Whelan, T. (1999). Decision-making in the physician-patient encounter: Revisiting the shared treatment decision-making model. *Social Science & Medicine*, 49, 651-661.
- Chassin, M.R., Kosecoff, J., Park, R.E., Winslow, C.M., Kahn, K.L., Merrick, N. J., Keeseey, J., Fink, A., Solomon, D. H., & Brook, R. H. (1987). Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures, *Journal of the Medical Association*, 258(18), 2533-2537.
- Chassin, M.R., Brook, R.H., & Park, R.E. (1986). Variations in the use of medical and surgical services by the Medicare population. *New England Journal of Medicine*, 314, 285-290.
- Chinburapa, V., Larson, L.N., Brucks, M., Draugalis, J., Bootman, J.L., & Puto, C.P. (1993). Physician prescribing decisions: The effects of situational involvement and task complexity on information acquisition and decision making. *Social Science in Medicine*, 36 (11), 1473–1482.
- Christensen-Szalanski, J.J.J. (1980). A further examination of the selection of problem-solving strategies: The effects of deadlines and analytic aptitudes. *Organizational Behavior and Human Performance*, 25(1), 107-122.
- Cohen, L.J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Science*, 4, 317-370.
- Cooksey, R.W. (1996). *Judgment Analysis: Theory, Methods, and Applications*. Academic Press: San Diego, CA.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. (1999). How Good Are Simple Heuristics? In: G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that makes us smart*. Oxford, New York: University Press.
- Davidson, J.R., Brundage, M.D., & Feldman-Stewart, D. (1999). Lung cancer treatment decisions: Patients' desires for participation and information. *Psycho-Oncology*, 8(6), 511-520.

- Davison, B.J., Degner, L.F., & Morgan, T.R. (1995). Information and decision-making preferences of men with prostate cancer. *Oncology Nurses Forum*, 22(1), 1401-1408.
- Dawes, R.M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81(2), 95-106.
- Deber, R. (1994). Physicians in health care management: 7. The patient-physician partnership: Changing roles and the desire for information. *Canadian Medical Association Journal*, 151(2), 171-176.
- Degner, L.F., & Sloan, J.A. (1992). Decision making during serious illness: What role do patients really want to play? *Journal of Clinical Epidemiology*, 45(9), 941-950.
- Degner, L.F., Kristjanson, L.J., & Bowman, D. (1997). Information needs and decisional preferences in women with breast cancer. *Journal of the American Medical Association*, 277(18), 1485-1492.
- Detmer, D.E., Fryback, D.G., & Gassner, K. (1978). Heuristics and biases in medical decision-making. *Journal of Medical Education*, 53(8), 682-683.
- Dhami, M.K. (2003). Psychological models of professional decision-making. *Psychological Science*, 14(2), 175-180.
- Dhami, M.K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making*, 14(2), 141-168.
- Dhami, M.K., & Harries, C. (2001). Fast and frugal versus regression models of human judgment. *Thinking and Reasoning*, 7(1), 5-27.
- Dockser Marcus, A. (2004). Price becomes factor in cancer treatment. Costly 'targeted' drugs extend lives, but confront patients with wrenching choices. *The Wall Street Journal*, September 7.
- Dorans, N., & Drasgow, F. (1978). Alternative weighting schemes for linear prediction. *Organizational Behavior and Human Performance*, 21(3), 316-345.
- Ebbesen, E.B., & Konecni, V.J. (1980). On the external validity of decision making research: What do we know about decisions in the real world? In: T.S. Wallsten (Ed.), *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ: Lawrence Erlbaum Association.
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Eddy, D.M. (1990). Clinical decision making: From theory to practice. Anatomy of a decision. *Journal of the American Medical Association*, 263, 441-443.
- Edland, A. (1989). On cognitive processes under time stress: A selective review of the literature on time stress and related stress. Stockholm: Department of Psychology, University Stockholm.
- Edland, A. (1994). Time pressure and the application of decision rules: Choices and judgments among multiattribute alternatives. *Scandinavian Journal of Psychology*, 35(3), 281 - 291.

- Einhorn, H.J., & Hogarth, R.M. (1981). Rationality and the sanctity of competence. *The Behavioral and Brain Science*, 4, 334-335.
- Einhorn, H.J., & Hogarth, R.M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13(2), 171-192.
- Einhorn, H.J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 73(3), 221-230.
- Einhorn, H.J. (1971). Use of nonlinear, noncompensatory models as function of task and amount of information, *Organizational Behavior and Human Performance*, 6(1), 1-27.
- Einhorn, H.J., Kleinmuntz, D.N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 8(5), 465-485.
- Elstein, A. S. (1988). Cognitive processes in clinical inference and decision making. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference and judgment in clinical psychology*. New York: The Free Press.
- Emanuel, E.J., & Emanuel, L.L. (1992). Four models of the physician-patient relationship. *Journal of the American Medical Association*, 267, 2221-2226.
- Engelhardt, K. (2006). Patientenzentrierte Medizin im Spannungsfeld der Interessen. *Schleswig – Holsteinisches Ärzteblatt*, 7, 45-48.
- Evans, J.St.B.T, Harries, C., Dennis, L., & Dean, J. (1995). General practitioners' tacit and stated policies in the prescription of lipid lowering agents. *British Journal of General Practice*, 45, 15-18.
- Fallowfield, L., & Jenkins, V. (1999). Effective communication skills are the key to good cancer care. *European Journal of Cancer*, 35(11), 1592-1597.
- Field, A (2005). *Discovering statistics using SPSS*. London: Sage Publication.
- Fischer, J.E., Steiner, F., Zucol, F., Berger, C., Martignon, L., Bossart, W. Altwegg, M., & Nadal, D. (2002). Use of simple heuristics to target macrolide prescription in children with community-acquired pneumonia. *Archives of Pediatrics and Adolescent Medicine*, 156, 1005-1008.
- Ford, J.K., Schmitt, N., Schechtman, S. L. Hults, B.M., & Doherty, M.L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Decision processes*, 4(3), 75-117.
- Forster, M., Martignon, L., Masanori, T., Takezawa, M., Vitouch, O., & Gigerenzer, G. (2002). Simple heuristics versus complex predictive instruments: Which is better and why? Unpublished manuscript.
- Fox, J. (1980). Making decisions under the influence of memory. *Psychological Review*, 87(2), 190-211.
- Fraser, J.M., Smith, P.J., & Smith, J.W. (1992). A catalog of errors. *International Journal of Man-Machine Studies*, 37(3), 265-307.
- Gattellari, M., Butow, P.N., Tattersall, M.H.N., Dunn, S.M., & McLeod, C.A. (1999). Misunderstanding in cancer patients: Why shoot the messenger. *Annals of Oncology*, 10(1), 39-46.

- Garcia-Retamero, R., & Hoffrage, U. (2006). How causal knowledge simplifies decision-making. *Minds and Machines*, 16 (3), 365-380.
- Gay, J.R., & Sax Jacobs, B.J. (1983). *The technology explosion in medical science: Implication for the health care industry and the public (1981-2001)*. New York: SP Medical and Scientific Books.
- Geman, S.E., Bienenstock, E., & Doursat, R. (1992). Neural network and the bias/variance dilemma. *Neural Computation*, 4, 1-58.
- Gigerenzer, G., Czerlinski, U., & Martignon, L. (1999). How good are fast and frugal heuristics? In: J.Shanteau, B.Mellers, and D. Schum (Eds.), *Decision Science and Technology: Reflections on the contribution of Ward Edwards*. Norwell, MA: Kluwer Academic Publishers.
- Gigerenzer, G. (2006). Bounded and Rational. In R.J. Stainton (Ed.), *Contemporary debates in cognitive science*. Oxford, UK: Blackwell.
- Gigerenzer, G., & Kurzenhäuser, S. (2005) Fast and Frugal heuristics in Medical Decision Making. In: R. Bibace, J.D. Laird, K.L. Noller & J. Valsiner (Eds.), *Science and medicine in dialogue: Thinking through Particulars and Universals* . Westport, CT: Praeger.
- Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.
- Gigerenzer, G., & Todd, P. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that makes us smart*. New York: Oxford University Press.
- Gigerenzer, G., & Goldstein, D. (1999). Betting on a good reason: The Take The Best heuristic. In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that makes us smart*. New York: Oxford University Press.
- Gigerenzer, G., Todd, P., & the ABC Research Group (1999). *Simple heuristics that makes us smart*. New York: Oxford University Press.
- Gigerenzer, G., Hoffrage, U. & Ebert, A. (1998). AIDS counseling for low-risk clients. *Aids Care*. 10 (2), 197-211
- Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650-669.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684-704.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98(4), 506-528.
- Gillis, J.S., Lipkin, J.O., & Moran, T.J. (1981). Drug therapy decisions, a social judgment analysis. *Journal of Nervous and Mental Disease*, 169(7), 439-447.
- Goldstein, D.B., Tate, S.K., & Sisodiya, S.M. (2003). Pharmacogenetics goes genomic. *Nature Review Genetics*, 4, 937-947.

- Gorman, C.D., Clover, W.H., & Dohert, M.E. (1978) Can we learn anything about interviewing real people from 'interviews' of paper people? Two studies of the external validity of a paradigm. *Organizational Behavior and Human Performance*, 22, 165-192.
- Greenhalgh, T., Kostopoulou, O., & Harries, C. (2004). Making decisions about benefits and harms of medicine. *British Medical Journal*, 329, 47-50.
- Hammond, J.S., Keeney, R.L., & Raiffa, H. (1998). Even swaps: A rational method for making trade-offs. *Harvard Business Review*, 76 (2), 137-150.
- Hammond, K.R. (1996). Upon reflection. *Thinking and Reasoning*, 2(3), 239-248.
- Hammond, K.R. (1971). Computer graphics as an aid of learning. *Science*, 172, 903-908.
- Harries, C., & Gilhooly, K. (2003). Identifying occupational therapists' referral priorities in community health. *Occupational Therapy International*, 10, 150-164.
- Harries, C., Evans, J.St.B.T, Dennis, L., & Dean, J. (1996). A clinical judgment analysis of prescribing decisions in general practice. *Le Travail Humain*, 59(1), 87-111.
- Henman, M.J., Butow, P.N., Brown, R.F., Boyle, F., & Tattersall, M.H. (2002). Lay constructions of decision-making in cancer. *Psycho-Oncology*, 11(4), 295-306.
- Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57(2), 116-131.
- Hoffrage, U., & Gigerenzer, G. (1996). The impact of information representation on Bayesian reasoning. In: G. Cottrell (Eds.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- Holtzman, N.A. (2003). Clinical Utility of Pharmacogenetics and Pharmacogenomics. In: Mark A. Rothstein (Eds.), *Pharmacogenomics: Social, ethical, and clinical dimensions*. New Jersey: John Wiley & Son, Inc.
- Iscoc, N.A., Goel, V., Wu, K., Fehringer, G., Holowaty, E.J., & Naylor, C.D. (1994). Variation in breast cancer surgery in Ontario. *Canadian Medical Association Journal*, 150(3), 345-352.
- Jenkins, V., Fallowfield, L., & Saul, J. (2001). Information needs of patients with cancer: Results from a large study in UK cancer centres. *British Journal of Cancer*, 84(1), 48-51.
- Johnson, E.J., & Meyer, R.J. (1984). Compensatory choice models of noncompensatory processes: The effect of varying context. *Journal of Consumer Research*, 11, 528-541.
- Jones. P.A., & Laird, P.W. (1999). Cancer-epigenetics comes of age. *Nature Genetics*, 21, 163-167.
- Jones, T.V., Gerrity, M.S., & Earp, J. (1990). Written case simulations: Do they predict physicians' behavior? *Journal of Clinical Epidemiology*, 43(8), 805-815.
- Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice-Hall.

- Kahneman, D., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Katz, S.J., Charles, C., Lomas, J. & Welch, H.G. (1997). Physician relations in Canada: Shooting inward as the circle closed. *Journal of Health Politics, Policy and Law*, 22(6), 1413-1431.
- Kee, F., Jenkin, J., McIlwaine, S., Patterson, C., Harper, S. & Shields, M. (2003). Fast and frugal models of clinical judgment in novice and experts physicians. *Medical Decision Making*, 7, 293-300.
- Kirwan, J.R., Chaput de Saintonge, D.M., Joyce, C.R.B., & Currey, H.L.F. (1984). Clinical judgment in rheumatoid arthritis, III. British rheumatologists' judgments of change in response to therapy. *Annals of Rheumatic Diseases*, 43(5), 686-694.
- Klayman, J. (1985). Children's decision strategy and their adaptation to task characteristics. *Organizational Behavior and Human Decision Processes*, 35(2), 179-201.
- Klein, G. (2001). The Fiction of Optimization. In: G. Gigerenzer and R. Selten (Eds.), *Bounded rationality - The Adaptive Toolbox*. Boston: MIT Press.
- Kurzenhäuser, S. (2003). Natural frequencies in medical risk communication: Application of a simple mental tool to improve statistical thinking in physicians and patients. Free University Berlin: Dissertation.
- Lam, W.W., Fielding, R., Chan, M., Chow, L. & Or, A. (2005). Gambling with your life: The process of breast cancer treatment decision making in Chinese women. *Psycho-Oncology*, 14(1), 1-15.
- Lazarou, J., Pomeranz, B.H., & Corey, P.N. (1998). Incidence of adverse drug reactions in hospitalized patients. *Journal of the American Medical Association*, 279(15), 1200-1205.
- Leape, L.L., Hilborn, L.H., Park, R.E., Bernstein, S. J., Kamberg, C. J., Sherwood, M., & Brook, R.H. (1993). The appropriateness of use of coronary artery bypass graft surgery in New York State. *Journal of the American Medical Association*, 269(6), 753-760.
- Levine, M.N., Gafni, A., Markham, B., & MacFarlane, D. (1992). A bedside decision instrument to elicit a patient's preference concerning adjuvant chemotherapy for breast cancer. *Annals of Internal Medicine*, 11(7), 53-58.
- Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, 26(2), 149-171.
- Lomas, J. (1990). Quality assurance and effectiveness in health care: An overview (editorial). *International Journal for Quality in Health Care*, 2, 5-12.
- Lomas, J., Anderson, G.M., Domnick-Pierre, K., Vayda, E., Enkin, M.W., & Hannah, W.J. (1989). Do practice guidelines guide practice? The effect of a consensus statement on the practice of physician. *New England Journal of Medicine*, 321(19), 1306-1311.

- Lomas, J., & Contandriopoulos, A.P. (1994). Regulating limits to medicine: Towards harmony in public- and self-regulation. In: R.G. Evans, M.L. Barer & T.R. Marmor (Eds.), *Why are some people healthy and others not? The determinants of the health of populations*. New York: Adline De Gruyter.
- Louviere, J., Hensher, D., & Swait, J. (2000). *Stated choice methods: Analysis and application*. Cambridge: Cambridge University Press.
- Martignon, L. (2001). Comparing fast and frugal heuristics and optimal models. In: G. Gigerenzer, & R. Selten (Eds.), *Bounded rationality - The Adaptive Toolbox*. Boston: MIT Press.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52 (1), 29-71.
- McAllister, D. W., Mitchell, T.R., & Beach, L.R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance*, 24, 228-244.
- McVea, K.L., Minier, W.C., & Johnson Palensky, J.E. (2001). Low-income women with early-stage breast cancer: Physician and patient decision-making styles. *Psycho-Oncology*, 10(2), 137-146.
- Meredith, C., Symonds, P., Webster, L., Lamont, D., Pyper, E., Gillis, C.R., & Fallowfield, L. (1996). Information needs of cancer patients in West Scotland: Cross sectional survey of patients' view. *British Medical Journal*, 313, 724-726.
- Miller, G.A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-93.
- National Cancer Alliance (1996). *Patient – centred cancer services? What patients say*. Oxford: National Cancer Alliance.
- Nelson, W., Stefanek, M., Peters, E., & McPaul, K. D. (2005). Basic and applied decision making in cancer control. *Health Psychology*, 24 (4), 3-8.
- Olshavsky, R.W. (1979). Task complexity and contingent processing in decision making: A replication and extension. *Organizational Behavior and Human Performance*, 24(3), 300-316.
- Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceeding of the National Academy of Science of the United States of America*, 102 (24), 8776 – 8780.
- Onken, J., Hastie, R., & Revelle, W. (1985). Individual differences in the use of simplification strategies in a complex decision-making task. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 14-27.
- Payne, J.W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16(2), 366-387.
- Payne, J.W., Bettman, J.R., James, R., & Johnson, E.J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14(3), 534-552.

- Payne, J.W., & Braunstein, M.L. (1978). Risky choice: An examination of information acquisition behavior. *Memory & Cognition*, 6(5), 554-561.
- Payne, J.W., Bettman, J.R., & Johnson, E.J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.
- Pirmohamed, M., James, S., Meakin, S., Green, C., Scott, A.K., Walley, T.J., Farrar, K., Park, B.K., Breckenridge, A.M. (2004). Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18 820 patients. *British Medical Journal*, 329, 15-19.
- Plous, S. (1993). *The Psychology of Judgment and Decision Making*. Philadelphia: Temple University Press.
- Poses, R.M., Cebul, R.D., Wigton, R.S., Centor, R.M., Collins, M., & Fleischli, G. (1992). Controlled trial using computerized feedback to improve physicians' diagnostic judgments. *Academic Medicine*, 67(5), 345-347.
- Poses, R.M., Bekes, C., Winkler, R.L., Scott, W.E., & Copare, F.J. (1990). Are two (inexperienced) heads better than one (experienced) head?: Averaging house officers' prognostic judgments for critically ill patients. *Archives of Internal Medicine*, 150(9), 1874-1878.
- Quill, T.E. (1983). Partnership in patient care: A contractual approach. *Annals of Internal Medicine*, 98(2), 228-234.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell. In: G. Gigerenzer, P. Todd, P. & the ABC Research Group (Eds.), *Simple heuristics that makes us smart*. Oxford New York: Oxford University Press.
- Roos, N.L. (1984). Hysterectomy: Variations in rates across small areas and across physicians' practices. *American Journal of Public Health*, 74, 327-335.
- Rothstein, H.G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes*, 37, 83-92.
- Sadée, W., & Dai, Z. (2005). Pharmacogenetics/genomics and personalized medicines. *Human Molecular Genetics*, 14(2), 207-214.
- Salkeld, G., Solomon, M. & Butow, P. (2004). A matter of trust – patient's views on decision making in colorectal cancer. *Health Expectations*, 7, 104-114.
- Schmidt, F.L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 31(3), 699-714.
- Scholl, W. (2003). Grundkonzepte der Organisationsforschung. In: H. Schuler (Ed.), *Lehrbuch der Organisationspsychologie*. Bern: Huber.
- Scholl, W. (2004). *Innovation und Information. Wie in Unternehmen neues Wissen produziert wird*. Göttingen: Hogrefe.

- Schwartz, W.B., Gorry, G.A., Kassirer, J.P., & Essig, A. (1973). Decision analysis and clinical judgment. *The American Journal of Medicine*, 55(4), 459-472.
- Schwartz, S. & Griffin, T. (1986). *Medical Thinking: The Psychology of Medical Judgment and Decision Making*. New York: Springer.
- Scott, A. (2002). Identifying and analyzing dominate preferences in discrete choice experiments: An application in health care. *Journal of Economic Psychology*, 23(3), 383-398.
- Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 130(3), 380-400.
- Shields, M.D. (1980). Some effects on information load on search patterns used to analyze performance reports. *Accounting, Organizations and Society*, 5(4), 429-442.
- Simon, H.A. (1957). *Models of man: Social and rational*. New York: John Wiley & Sons, Inc.
- Simon, H.A. (1956). Rational choice and the structure of environment. *Psychological Review*, 63(2), 129-163.
- Slaytor, E.K., & Ward, J.E. (1998). How risks of breast cancer and benefits of screening are communicated to women – Analysis of 58 pamphlets. *British Medical Journal*, 317, 263-264.
- Slovic, P., Rorer, L.G., & Hoffman, P.J. (1971). Analyzing the use of diagnostic signs. *Investigative Radiology*, 6(1), 18-26.
- Smith, L., & Gilhooly, K. (2006). Regression versus fast and frugal models of decision-making: The case of prescribing for depression. *Applied Cognitive Psychology*, 20(2), 265-274.
- Smith, L., Gilhooly, K., & Walker, A. (2003). Factors influencing prescribing decisions in the treatment of depression: A social judgment theory approach. *Applied Cognitive Psychology*, 17(1), 51-63.
- Sox, H.C., Blatt, M.A., Higgins, M.C., & Marton, K.I., (1988). *Medical Decision Making*. Woburn: Butterworth & Heinemann.
- Speroff, T., Connors, A.F., & Dawson, N.V. (1989). Lens model analysis of hemodynamic status in the critically ill. *Medical Decision Making*, 9(4), 243-252.
- Staelin, R., & Payne, J.W. (1976). Studies of information seeking behavior of consumers. In: J.S. Carroll & J.W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, New Jersey: Erlbaum.
- Statistisches Bundesamt Deutschland (2006). *Gesundheitsbereichterstattung des Bundes. Gesundheitsausgaben in Deutschland in Millionen*. Received from: http://www.gbe-bund.de/oowa921install/servlet/oowa/aw92/dboowasys921.xwdevkit/xwd_init?gbe.isgbetol/xs_start_neu/371511452/21318197
- Stegina, S.K. & Ochipinti, S. (2002). Decision making about treatment of hypothetical prostate cancer: Is deferring a decision an expert-opinion heuristic? *Journal of Psychosocial Oncology*, 20 (3), 69-84.
- Stewart, T. (1988). Judgment analysis: Procedures. In: B. Brehmer & C.R.B. Joyce (Eds.), *Human Judgment: The SJT View*. Amsterdam: Elsevier.

- Strauss, G., Chassin, M., & Lock, J. (1995). Can experts agree when to hospitalize adolescents? *Journal of the American Academy of Child and Adolescent Psychiatry*, 4, 418-424.
- Sundstrom, G.A. (1987). Information search and decision making: The effects of information display. *Acta Psychologica*, 65(2), 165-179.
- Sutherland, H.J., Llewellyn-Thomas, H.A., Lockwood, G.A., Trichtler, D.L., & Till, J.E. (1992). Cancer patients: Their desire for information and participation in treatment decisions. *Journal of the Royal Society of Medicine*, 82(5), 260-263.
- Tabachnick, B.G. & Fidell, L.S. (1996). *Using Multivariate Statistics*. Harper Collins: New York.
- Üstün, T.B., & Sartorius, N. (1995) *Mental illness in general health care*. Wiley: Chichester.
- Veatch, R.M. (1972). Models of ethical medicine in a revolutionary age. *Hastings Center Report*, 2, 5.
- Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83(2), 213-217.
- Wallsten, T.S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 361-384.
- Wennberg, J.E. (1990). Better policy to promote the evaluative clinical sciences. *International Journal for Quality in Health Care*, 2, 21-29.
- Wertman, B.G., Sostrin, S.V., Pavlova, Z., & Lundberg, G.D. (1980). Why do physicians order laboratory tests? A study of laboratory test request and use patterns. *Journal of the American Medical Association*, 243(20), 2080-2082.
- Wigton, R.S., Hoellerich, V.L., & Patil, K.D. (1986). How physicians use clinical information in diagnosing pulmonary embolism: An application of conjoint analysis. *Medical Decision Making*, 6(1), 2-11.
- Wigton, R.S. (1987). The use of computer simulation in teaching clinical diagnosis. *Computer Methods and Programs in Biomedicine*, 25(2), 111-114.
- Wigton, R.S. (1996). Social judgment theory and medical judgment. *Thinking and Reasoning*, 2(2), 175-190.
- Wright, P. (1974). The harassed decision maker: Time pressure, distractions, and the use of evidence. *Journal of Applied Psychology*, 59(5), 555-561.